

Lecture Notes on APPLIED EXPERIMENTAL DESIGNS FOR AGRICULTURAL RESEARCH

VISIT...

LANZAROTE
Caliente.COM

APPLIED EXPERIMENTAL DESIGNS FOR AGRICULTURAL RESEARCH

Efren C. Altoveros¹

There are two keywords in the topic: experiment and design. An experiment is the act of conducting a controlled test or investigation. The word controlled means most, if not all, of the conditions that happened or were used in the experiment are known or regulated. In the field of agriculture, an experimental research is conducted to answer a particular question or solve a particular problem. In experimental research, different kinds or levels of a particular factor or several factors are evaluated.

The second keyword is design which may mean arrangement. In agricultural research, proper design is important because we want to establish or find the true results without any doubt in mind. With improper or wrong design, results may not be convincing or reliable.

Research means systematic investigation to establish facts. It is systematic because all the activities are planned and executed based on rules so everything can be repeated. The term established facts indicates that research is done to prove something that has been done before.

The aim in conducting research is find out if significant differences exist among the levels or kinds of treatments (or factor combinations). This is accomplished by using the technique of ANALYSIS OF VARIANCE. Note that the treatment means are, in most cases, different from one another. The question is: is the difference significant enough to conclude that one treatment is better (or bigger, smaller, etc.) than the other. The difference is considered significant if the variation among treatments is proportionally larger than the variation due to unexplained error. Analysis of variance uses two basic estimates: mean and deviation from the mean.

Before going into the computation of mean and deviation, some notations need to be understood and remembered. Given a set of 10 values assigned to a variable X and to variable Y:

$X_1 = 4$	$Y_1 = 7$
$X_2 = 3$	$Y_2 = 8$
$X_3 = 5$	$Y_3 = 6$
$X_4 = 2$	$Y_4 = 5$
$X_5 = 6$	$Y_5 = 7$
$X_6 = 7$	$Y_6 = 9$
$X_7 = 3$	$Y_7 = 5$
$X_8 = 6$	$Y_8 = 9$
$X_9 = 4$	$Y_9 = 8$
$X_{10} = 5$	$Y_{10} = 6$

1. Population is a group of individuals, objects, or things possessing at least one common character or trait not possessed by any other individuals, objects or things. Examples are human population, plant population, population of one country, bacterial population, students enrolled in a particular university, books in a library, etc.

In studying statistics, it is not always possible to deal with populations because of their enormous size. For example, if we want to know the average height of the human population, it is impossible to measure each and every human being. If we want to know the average yield of wheat in a particular country, we cannot weigh all the what harvested in that country. In this case, we have to deal with a sample where proper statistical analysis can be done.

¹ Training Consultant and Lecturer, Calamba City, Philippines.
E-mail: efren@altoveros.net; ealtoveros@yahoo.com

2. Sample is a representative group taken at random from a population. When a sample is properly taken, the statistics from that sample can be applied to the population. The 10 values given above is an example of a sample.
3. **X** (or **Y**) is a symbol used to designate a variable. A variable is something that changes in value. In contrast, a constant is something that remains the same. For example, if there are 10 children in the group, they can be designated $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$, and X_{10} whose ages are 4, 3, 5, 2, 6, 7, 3, 6, 4, and 5 years, respectively.
4. Σ is the symbol to designate summation or sum. In the example, the sum of X_i where i goes from 1 to 10 is:

$$\begin{aligned}\Sigma X_i &= X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} \\ &= 4 + 3 + 5 + 2 + 6 + 7 + 3 + 6 + 4 + 5 = \boxed{45}\end{aligned}$$

5. **N** refers to the number of observations in a population.
6. μ refers to the population mean. It is the average value of all observations in a population.

$$\mu = \frac{\Sigma X_i}{N} = \frac{X_1 + X_2 + \dots + X_N}{N}$$

7. **n** refers to the number of observations in a sample. In the example, **n = 10**.
8. $\bar{X}$ refers to sample mean. It is the average value of all observations in a sample. By formula:

$$\begin{aligned}\bar{X} &= \frac{\Sigma X_i}{n} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10}}{10} \\ &= \frac{4 + 3 + 5 + 2 + 6 + 7 + 3 + 6 + 4 + 5}{10} = \frac{45}{10} = \boxed{4.5}\end{aligned}$$

9. ΣX^2 refers to the sum of the squares of individual observation. In the formula, each value is squared first and then added.

$$\begin{aligned}\Sigma X^2 &= X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2 + X_6^2 + X_7^2 + X_8^2 + X_9^2 + X_{10}^2 \\ &= 4^2 + 3^2 + 5^2 + 2^2 + 6^2 + 7^2 + 3^2 + 6^2 + 4^2 + 5^2 \\ &= 16 + 9 + 25 + 4 + 16 + 49 + 9 + 36 + 16 + 25 = \boxed{225}\end{aligned}$$

10. $(\Sigma X)^2$ refers to the square of the sum of individual observations. In the formula, all the values are added first and the total is squared.

$$\begin{aligned}(\Sigma X)^2 &= (X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10})^2 \\ &= (4 + 3 + 5 + 2 + 6 + 7 + 3 + 6 + 4 + 5)^2 = 45^2 = \boxed{2025}\end{aligned}$$

11. ΣXY refers to the sum of the **product of X & Y**, that is, the combinations of X & Y are multiplied first before adding.

$$\begin{aligned}\Sigma XY &= X_1Y_1 + X_1Y_2 + \dots + X_iY_j \\ &= (4 \times 7) + (3 \times 8) + \dots + (5 \times 6) = 28 + 24 + \dots + 30 = \boxed{328}\end{aligned}$$

12. $\Sigma X \Sigma Y$ refers to the product of the **sum of X** and **sum of Y**, that is, all X's are added first as well as all Y's and the sums are multiplied.

$$\begin{aligned}\Sigma X \Sigma Y &= (X_1 + X_2 + \dots + X_i) \times (Y_1 + Y_2 + \dots + Y_j) \\ &= (4 + 3 + \dots + 5) \times (7 + 8 + \dots + 6) = 45 \times 70 = \boxed{3150}\end{aligned}$$

13. The sum of individual numbers each multiplied by a constant is equal to the over-all sum of the numbers multiplied by a constant.

$$= \Sigma[(X_1 \times c) + (X_2 \times c) + (X_3 \times c) + (X_4 \times c) + (X_5 \times c) + (X_6 \times c) + (X_7 \times c) + (X_8 \times c) + (X_9 \times c) + (X_{10} \times c)]$$

$$= \Sigma(X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10}) \times c$$

If each number in the sample is to be multiplied by 10,

$$4 \times 10 + 3 \times 10 + 5 \times 10 + 2 \times 10 + 6 \times 10 + 7 \times 10 + 3 \times 10 + 6 \times 10 + 4 \times 10 + 5 \times 10$$

$$= (4 + 3 + 5 + 2 + 6 + 7 + 3 + 6 + 4 + 5) \times 10$$

$$40 + 30 + 50 + 20 + 60 + 70 + 30 + 60 + 40 + 50 = (45) \times 10 = \boxed{450}$$

14. The sum of individual numbers each divided by a constant is equal to the over-all sum of the numbers divided by a constant.

$$\Sigma \left[\frac{X_1}{c} + \frac{X_2}{c} + \frac{X_3}{c} + \frac{X_4}{c} + \frac{X_5}{c} + \frac{X_6}{c} + \frac{X_7}{c} + \frac{X_8}{c} + \frac{X_9}{c} + \frac{X_{10}}{c} \right]$$

$$= \frac{\Sigma(X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10})}{c}$$

If each value in the sample data is divided by a constant value 10,

$$\frac{4}{10} + \frac{3}{10} + \frac{5}{10} + \frac{2}{10} + \frac{6}{10} + \frac{7}{10} + \frac{3}{10} + \frac{6}{10} + \frac{4}{10} + \frac{5}{10}$$

$$= \frac{(4 + 3 + 5 + 2 + 6 + 7 + 3 + 6 + 4 + 5)}{10}$$

$$(0.4 + 0.3 + 0.5 + 0.2 + 0.6 + 0.7 + 0.3 + 0.6 + 0.4 + 0.5) = 45/10 = \boxed{4.5}$$

Deviation is the difference between an observation and the mean. Deviation is positive if the observation value is larger than the mean and negative if it is smaller than the mean. The sum of the deviation of all observations from the mean is zero. In the example, the deviations will be:

Sample	Mean	Deviation
4	4.5	-0.5
3	4.5	-1.5
5	4.5	0.5
2	4.5	-2.5
6	4.5	1.5
7	4.5	2.5
3	4.5	-1.5
6	4.5	1.5
4	4.5	-0.5
5	4.5	0.5
		0.0

Since the total deviation is zero, the use of deviation alone is not advisable in determining how much variation is there between the observations and the mean. There is a need to make all the deviations positive in value. One of the techniques commonly used is the Method of Least Squares. In this method, the deviation value is first squared (hence all values will be positive) and then added to become Sum of Squares.

Sample	Mean	Deviation	Dev ²
4	4.5	-0.5	0.25
3	4.5	-1.5	2.25
5	4.5	+0.5	0.25

2	4.5	-2.5	6.25
6	4.5	+1.5	2.25
7	4.5	+2.5	6.25
3	4.5	-1.5	2.25
6	4.5	+1.5	2.25
4	4.5	-0.5	0.25
5	4.5	+0.5	0.25
			22.50

When the Sum of Squares is divided by the number of observations, the value obtained is known as Variance (designated by the symbol σ^2 when dealing with a population and by s^2 when dealing with a sample). By definition, Variance is the average of the squared deviation from the mean. By formula:

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N} \text{ when dealing with a population, and}$$

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{(n - 1)} \text{ when dealing with a sample}$$

In the sample data, variance is computed as:

$$s^2 = \frac{(4-4.5)^2 + (3-4.5)^2 + (5-4.5)^2 + (2-4.5)^2 + (6-4.5)^2 + (7-4.5)^2 + (3-4.5)^2 + (6-4.5)^2 + (4-4.5)^2 + (5-4.5)^2}{10 - 1}$$

$$= \frac{(-0.5)^2 + (-1.5)^2 + (0.5)^2 + (-2.5)^2 + (1.5)^2 + (2.5)^2 + (-1.5)^2 + (1.5)^2 + (-0.5)^2 + (0.5)^2}{9}$$

$$= \frac{(0.25) + (2.25) + (0.25) + (6.25) + (2.25) + (6.25) + (2.25) + (2.25) + (0.25) + (0.25)}{9} = \mathbf{22.50}$$

The $(n - 1)$ refers to the degree of freedom or df. By definition, degree of freedom is the number of samples values needed to get the total given the mean value. In statistics or experimental designs, df is used as divisor of the sum of squares in obtaining variance because we are dealing with sample and not the whole population. This will make the value of the sample variance closer to the actual value of the population variance.

Although these are the original variance formula by definition, they are difficult to use when using with means with recurring decimals, e.g. 3.333333..., 2.677777..., 3.57142857142857.... When these means are subtracted from the individual observations, the deviations are not accurate because of rounding off of decimals that will lead to inaccurate answer. Consider a set of 13 values below:

X_1	4
X_2	3
X_3	5
X_4	2
X_5	6
X_6	7
X_7	3
X_8	6
X_9	4
X_{10}	5
X_{11}	5
X_{12}	4
X_{13}	7
TOTAL	61
MEAN	4.69230769230769

When the deviations and squared deviations are computed, the values will be:

X ₁	4	-0.69230769230769	0.47928994082840
X ₂	3	-1.69230769230769	2.86390532544379
X ₃	5	0.30769230769230	0.09467455621301
X ₄	2	-2.69230769230769	7.24852071005917
X ₅	6	1.30769230769231	1.71005917159763
X ₆	7	2.30769230769231	5.32544378698225
X ₇	3	-1.69230769230769	2.86390532544379
X ₈	6	1.30769230769231	1.71005917159763
X ₉	4	-0.69230769230769	0.47928994082840
X ₁₀	5	0.30769230769230	0.09467455621301
X ₁₁	5	0.30769230769230	0.09467455621301
X ₁₂	4	-0.69230769230769	0.47928994082840
X ₁₃	7	2.30769230769231	5.32544378698225
TOTAL	61		28.7692307692308
s ²			2.40

It can clearly be seen that manual computation (or use of calculator) of individual deviations will be difficult because the grand mean is not a whole number so the deviations (and the squares) are not whole numbers too. The common (and recommended) procedure is to round-off the decimals to the nearest 2 decimal points. Sometimes, this may result to inaccurate final answer. To solve this problem or limitation, a machine formula for the sum of squares was derived to simplify computation. Through derivation, the original long formula can be converted into machine formula where:

$$\Sigma(X - \bar{X}) = \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{N}$$

To prove:

$$\begin{aligned}
 \Sigma(X - \bar{X}) &= \Sigma(X_i^2 - 2X\bar{X} + \bar{X}^2) \\
 &= \Sigma X_i^2 - \Sigma 2X\bar{X} + \Sigma \bar{X}^2 \\
 &= \Sigma X_i^2 - 2\Sigma X\bar{X} + \Sigma \bar{X}\bar{X} \\
 &= \Sigma X_i^2 - \frac{2\Sigma X\Sigma X}{N} + \bar{X}\Sigma \bar{X} \\
 &= \Sigma X_i^2 - \frac{2(\Sigma X)^2}{N} + \frac{\Sigma X\Sigma X}{N} \\
 &= \Sigma X_i^2 - \frac{2(\Sigma X)^2}{N} + \frac{(\Sigma X)^2}{N} \\
 &= \Sigma X_i^2 - \frac{(\Sigma X)^2}{N}
 \end{aligned}$$

Therefore, variance can be computed as:

$$\sigma^2 = \frac{\Sigma(X - \bar{X})}{N} = \frac{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{N}}{N} \text{ when dealing with a population, and}$$

$$s^2 = \frac{\Sigma(X - \bar{X})}{n - 1} = \frac{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}}{n - 1} \text{ when dealing with a sample}$$

Using the sample data, the sample variance can be computed as:

$$s^2 = \frac{4^2+3^2+5^2+2^2+6^2+7^2+3^2+6^2+4^2+5^2+5^2+4^2+7^2 - \frac{(4+3+5+2+6+7+3+6+4+5+5+4+7)^2}{13}}{13-1}$$

$$s^2 = \frac{16+9+25+4+36+49+9+36+16+25+25+16+49 - \frac{(61)^2}{13}}{12} = \frac{315 - 286.2307}{12} = \frac{28.7692}{12} = \boxed{2.40}$$

Standard Deviation (designated by the symbol σ when dealing with a population and by s when dealing with a sample) is the square root of Variance. It is used as a measure of average deviation from the mean. In the example, the standard deviation is $\sqrt{(2.40)} = \boxed{1.55}$

Normally, Mean should be expressed together with Standard Deviation to have a complete meaning. In the example, the mean should be expressed as $\boxed{4.49 \pm 1.55}$

On average, the range of values is from 3.14 to 6.24 ($4.69 - 1.55 = 3.14$; $4.69 + 1.55 = 6.24$). Note that the range will not be equal to the actual minimum and maximum values because there are few samples. As the number of samples increases, the range based on standard deviation will more or less be equal or very close to the range of values of the actual data.

Techniques in implementing a good experiment

As indicated in the discussion of the different experimental designs particularly field experiment, replication is needed to create experimental or unexplained error (variation) but proper blocking is done to make the unexplained error within the acceptable limit. As shown in several examples, too high unexplained error may make the experiment unacceptable while a very low unexplained error can cause problem in proper interpretation of results and recommendations.

Even if proper planning and blocking are accomplished, it does not guarantee that the results are reliable if the proper methods of carrying out the experiment in terms of cultural management, data gathering and harvesting are not followed.

Choosing a good experimental site

1. Slope. Generally, fertility gradients are more pronounced in sloping areas. The main reason is the movement of nutrients in the soil due to water when it rains and during flooding. Ideally, experiments should be conducted in areas with no slopes (level) but this not possible most of the time. If this not avoidable, proper blocking is needed.
 - A. If the fertility gradient is going in one direction, blocks with long and narrow plots should be made perpendicular to the gradient.
 - B. If the slope is going on both directions (north to south and east to west), or when fertility gradient is not uniform, make the block as square as possible. If there are numerous treatments or if the needed plot dimension is square that will make the block not nearly square, try using two (or even three) sub-blocks.
2. Areas used for experiments in previous croppings. When the area to be used for a future experiment has been used in a previous experiment, study the nature of the previous study to determine if it will have any direct or serious effect on the outcome of the new experiment. The common problem occurs when the area was used in a fertilizer experiment using different rates, or when different cultural management practices were used in the previous experiment that may have an impact on the new experiment. The variability of the area may be altered or reduced by planting one crop, usually a uniform variety, for at least one season before conducting the new experiment.

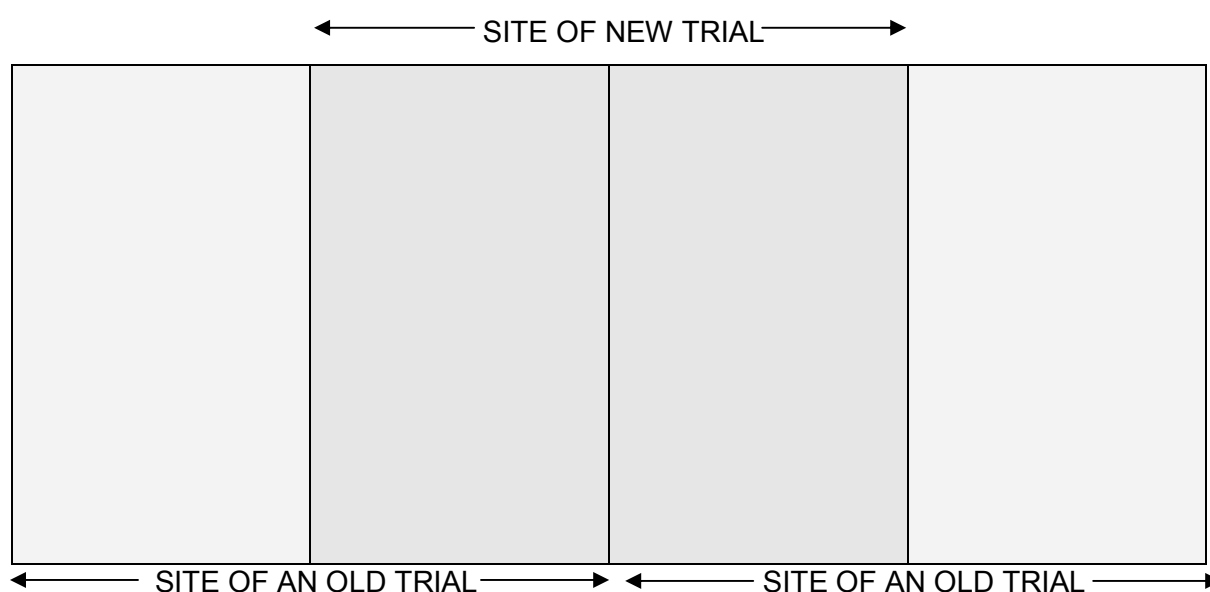
3. Graded areas. In some cases, the top soil from an elevated area is removed by grading to minimize the slope, thus exposing infertile subsoils. These types of soil should be avoided as much as possible. If it is not possible, proper determination of soil variability should be done and a suitable remedy be planned such as proper blocking or appropriate adjustments using covariance technique.
4. Presence of large trees, poles, and structures. Areas with surrounding permanent structures should be avoided, not only because of the direct effect of shading but also the nature of the soil near the structure. Soil movement during the construction of these structures may have occurred resulting to poor performance of the crop planted near the structure.
5. Unproductive sites. In conducting experiment, the aim is to have a productive crop so poor soil should be avoided unless it is the purpose of the experiment.

Solving problem of soil heterogeneity/variability

1. Plot size and shape. When the soil is suspected to be highly variable, small plot size with increased number of replications will minimize or reduce experimental error because the distance between any farthest points in each block will be shorter than when using large plots. As a result, the variability within each block (called intrablock error) is minimized.
2. Block size and shape. If the plot size cannot be reduced and it is suspected that the soil is highly variable with unknown direction, use of square-shaped blocks is recommended. The distance between any two farthest points in a square block is shorter than those in a long and narrow block.
3. Number of replications. If the soil is suspected to be highly variable, increasing the number of replications up, to a certain extent, will reduce experimental error. To be effective, this should be done together with decreasing plot size

Common mistakes in proper planning and proper lay-out of an experiment:

1. Failure to study the cropping history of the area to be used in the trial. One of the more common mistakes in planning an experiment is not taking into consideration what type of experiment was conducted in the area the previous year. Examples are:
 - a. The present trial is laid out in an area that covers part of the area used of one experiment before and part of the area of another experiment.



In the illustration, the site of the new trial will be conducted covering areas previously used in two different trials. This may result to uneven soil fertility and other conditions.

This can be partly corrected by proper land preparation but a better method is before the present experiment is conducted, plant one variety of one crop in the whole area without adding fertilizer. Let the crop grow to maturity and harvest (or plow under the crops in the case of legumes) during land preparation. This will more or less make the amount of residual elements uniform.

- b. A fertilizer study was done on the area and then a variety or pesticide trial will be conducted this time in the same area. Fertilizer trial uses different rates of a particular fertilizer and it is known that not all the nutrients are taken by the soil. This means plots given higher rates of fertilizer tend to retain higher amount of the nutrient compared to plots given lower rates. This makes the area not uniform in terms of soil fertility. If the land is not prepared adequately, areas with high residual amount of the element may give a better growing condition to the treatment planted there compared to areas with low residual amount.
2. Failure (or reluctance) to conduct soil analysis even if a fertilizer experiment is to be conducted in the area. Soil possesses natural elements needed by the plants for proper growth. It contains elements due to the fertilizer applied to the crops that are grown. It may also contain elements carried by wind, water and other factors. This makes the whole nutrient status of one area not uniform. When conducting a fertilizer experiment, it is necessary to conduct soil analysis (at least for the element(s) being studied) to know the initial amount of the element(s) before adding the treatment. This will help in the proper interpretation of final results.
3. Improper or uneven land preparation. Failure to prepare the land adequately may result to uneven plant growth and eventually poor yield or performance not because of the treatment applied.
4. Disregarding the effect of shading of surrounding trees or structures. There are plants that are highly sensitive to shading effect so proper blocking is necessary to take care of the problem in case portion of the area is partly shaded and there is nothing that can be done to move away from the shaded area.
5. Failure to consider the effect/influence of surrounding experiments. In planning a proper experiment, the existing experiments around the area should be considered. For example, if an experiment on organic farming is being conducted where spraying of pesticide is not allowed, a nearby experiment on insect control using insecticide will affect the outcome of the organic farming experiment since it is expected that the chemical residues will be carried by the wind (or water).

Cultural management mistakes committed while conducting an experiment

1. Improper or uneven application of fertilizer. In field experiments, fertilizer is applied application without measurement in case the trial is not a fertilizer experiment. Most of the time, application is not done properly which may result to uneven application of nutrients in the soil. This may affect the final result of the experiment and the proper effect of the treatment being studied may not be attained at all.
2. Uneven application of water. Similar to fertilizer application, improper or uneven water application is one of the common causes of error in field experiment. Low lying areas tend to receive more water that is retained at a much longer period of time. This may severely affect the plants in the area and thereby affecting the final result of the experiment.
3. Uneven (or sometimes unfinished) spraying of pesticide. This problem normally arises because of the tendency to finish or use the content of the spraying tank towards the end of spraying so the last portion of the area may receive too little or too much of the spray.
4. Unfinished cultural operation in a block. It is common practice of laborers to just stop weeding when it is time to go home even if weeding of the area in one block is not yet finished. This may result to an error in the data as plants in the portion not weeded will suffer from competition. Although the laborers cannot be prevented from leaving, the

people in-charge of the experiment should make it sure that weeding is not stated in a block if the whole process cannot be finished in the same day.

5. Failure to examine properly water source. If water is provided by irrigation canals, it is important to check what and where is the source and which areas the canals are passing through to be sure no unwanted fertilizer or chemical (not needed in the trial) goes with the water and be placed in the existing experiment.

Mistakes commonly committed during cultural management

1. Furrowing for row spacing. To reduce unexplained error, proper spacing between rows is necessary so that all plants are subjected to the same environmental condition. Improper alignment of rows may result to some plants having less plant-to-plant competition than others.
2. Selection of seedlings. The common practice in using seedlings as transplants is to select the best looking ones first and poor seedlings are leftovers. Also, seedlings at the outer side of the seedbeds are usually more vigorous than those at the middle so use of seedlings at the edge of the seedbed is not recommended.
3. Thinning. For most direct-seeded crops, a high seeding rate is used to ensure that enough seeds will germinate. Several days after germination, thinning (pulling out of excess plants) is done to maintain the number of plants to a constant number. This creates experimental error in the case of crops drilled in row. Aside from keeping the best-looking seedlings, plants have to be kept at the prescribed spacing. Sometimes, decision has to be made whether to keep the best looking seedlings or to maintain the required planting distance.
4. Transplanting. The common mistake in transplanting seedlings is the tendency to select the best seedlings such that the end plots usually receive the relatively poorer seedlings. This is minimized by assigning specific seedling trays per block.
5. Fertilizer application. Unless mechanical fertilizer applicator is used, there is always difficulty in applying fertilizer uniformly especially in large experimental area. This may be minimized by subdividing the area into smaller units before fertilizer is applied. Fertilizer per subunit can be measured and applied separately. Use of measuring device like tin cans proves to be useful in reducing error.
6. Seed mixtures and off-types. It is not easy to detect off types and seed mixtures when dealing with direct seeding or even with seedlings. Off types cannot be treated as normal plants because their performance is definitely affected not only by the treatments being evaluated but also by the genetic make-up different from the others. However, off types cannot be just ignored, pulled, or removed because by the time they are detected they could have affected surrounding plants through a competition effect different from the normal plants. Even if detection is early, removal will create missing hills that may affect surrounding plants.

Mistakes commonly committed during data gathering/harvest

1. Improper harvesting (not finishing the harvest in a block). When harvesting, it is a normal rule to finish harvesting of a block before going to the next block. Failure to harvest all entries in a block at the same time will result to bias in data gathering as the remaining plants may have the advantage of increasing its yield compared to the ones already harvested. Likewise, failure to complete harvesting of a block may result to bias to the plants left behind should unfavorable weather conditions such as heavy rain, strong wind, etc. occur before harvest is completed. Like in cultural operation, proper planning is necessary to ensure that all entries in a block are harvested at the same time.

There are situations that harvesting of selected entries is done ahead of the others. This is especially true on experiments involving varieties with different maturity. The early maturing varieties should be harvested ahead of others to prevent loss of fruit (or pods) if

not harvested as needed. A good example is field soybean. It is known that when soybeans reach full maturity, delay in harvesting may result to pod shattering hence loss of yield.

2. Improper instrument/measuring device used in data gathering. These include wrong balance (too big or too small capacity), or measuring device (too long or too short). As an example, it is not proper to use a 30-kg balance in getting average weight of potato tuber since the reading will not be accurate. Likewise, while a 1000-g balance is more accurate than a 10-kg balance, it is not appropriate to use it in weighing plot yield more than 1 kg since it will take more time to finish data gathering.

Another example is using a 30-cm ruler in measuring the diameter of green bean pods when a Vernier caliper would have been a better choice.

3. Delayed or staggered gathering of data. Like harvesting, delayed or staggered data gathering may give undue advantage (or disadvantage) for one entry over the other. The general rule is as much as possible, gather one set of data or observation for the whole experiment at the same time. If it is not possible to complete data gathering for the whole experiment, DO NOT start gathering data in a block that will not be finished in the same day.
4. Inadequate or improper sampling. In conducting experiment, adequate plot size is needed to get a good estimate of the data being gathered. However, there are situations where it is not possible (or practical) to harvest all plants so samples are taken to represent the whole plot. While this is acceptable, it is still much better if the whole plot is harvested. In case it is not, adequate sample size is necessary to get a good estimate of the real data. Very small sample may become bias since more often than not, samples are “selected” and not done at random. Because of this practice, the estimate is normally higher than what would have been the real data of the whole plot.
5. Tendency to select the best plant for sampling. When gathering data for some characters, plant samples are sometimes necessary because there is no need to measure all the plants in a plot. Examples are plant height, leaf length, number of seeds per pod, etc. In these cases, plant (or plant part) samples are taken to get an estimate of the values. In most cases, there is a tendency to select the best plants as samples making the values of the samples higher than what they should have been.
6. Entering data on loose paper. It is always recommended to use field book in gathering data and enter all information in the field book. Many researchers have the habit of entering data in loose sheets of paper and then transfer the data into the field book upon returning to office. This is not a good habit. as loose papers may be misplaced or lost along the way. As a general rule, a nice and clean data sheet is the worst data sheet.
7. Recalling data from memory. While some people have very sharp memory and can recall a lot of things easily, some people have not. As a rule, it is not advisable to just remember the values or observations and then enter them upon returning to office. A good researcher always carry a small field book whenever he goes to the field to record whatever observations he may see while visiting the field.
8. Estimation of data if they were not properly entered (or were not entered at all). When the memory fails to recall what should have been the values, or when the researcher did not notice that a certain value was not entered properly, estimation of the data is normally done. Again, this is not advisable as it causes error in the interpretation of final results. If the values cannot really be retrieved, it is better to apply missing value technique (or remove one entry or remove one block) when doing the final analysis.

PRINCIPLES OF ANALYSIS OF VARIANCE

When the word ANALYSIS is mentioned, it means identifying and dividing (or partitioning) the parts of a whole into meaningful components. A good example is blood analysis. Blood sample is taken from a person (or an animal) and then sent to the laboratory where the laboratory technician identifies what are the components of the blood and how much of each.

The same principle applies with Analysis of Variance. It basically means the TOTAL VARIANCE is divided (or partitioned) into meaningful components or parts to have a better idea how a particular object (person, plant, animal, etc.) reacts to a change in the conditions applied to it and how the environment plays a role in the expression of this reaction.

The total variance has two basic components:

- Explained variance - variance due to block and variance due to treatment or factor combinations.
- Unexplained variance - variance due to unknown or uncontrolled factors. This is also termed as experimental error.

To explain further the concept of explained and unexplained variances and how important they are in experimental designs, the following illustration might help. Consider a series of experiments involving growing of tomato in pots.

In Experiment #1, two tomato plants were grown in two large clay pots. The assumptions are:

- The two pots were planted to different varieties: Var 1 in pot #1 and Var 2 in pot #2.
- Both plants were grown side by side at the same time.
- Both pots contained exactly the same type of soil.
- Both plants were given the same amount of fertilizer and pesticide.
- Both plants were given the same cultural management practices such as cultivation, weed control, watering, etc.

After completing the harvest, the following yield data were gathered:

	<u>Var 1</u>	<u>Var 2</u>
Number of fruits	7	5
Yield	350 g	450 g

Since the growing conditions were the same for the two plants, it can be concluded that Var 2 gave a higher yield than Var 1, there were more fruits per plant in Var 1, and average fruit size in Var 2 was higher.

In Experiment #2, two tomato plants were grown in two large clay pots. The assumptions are:

- The two pots contained different types of soil: pot #1 had sandy loam while pot #2 has clay soil.
- Both pots were planted to the same variety Var 2.
- Both plants were grown side by side at the same time.
- Both plants were given the same amount of fertilizer and pesticide.
- Both plants were given the same cultural management practices such as cultivation, weed control, watering, etc.

After completing the harvest, the following yield data were gathered:

	<u>Sandy loam soil</u>	<u>Clay soil</u>
Total number of fruits	6	5
Total fruit yield	475 g	425 g

Since the two pots were grown to the same variety and given the same cultural management practices although the soil type differed, it can be concluded that the difference in yield was due to the type of soil and that sandy loam soil can give a higher yield than clay soil.

In Experiment #3, two tomato plants were grown in two large clay pots. The assumptions are:

- Both pots were planted to the same variety Var 2.
- Both plants were grown side by side at the same time.
- Both pots contain exactly the same type of soil.
- Both plants were given the same amount of fertilizer and pesticide.
- Both plants were given the same cultural management practices such as cultivation, weed control, watering, etc.

After completing the harvest, the following yield data were gathered:

	<u>Pot #1</u>	<u>Pot #2</u>
Total number of fruits	6	5
Total fruit yield	440 g	460 g

Note that in Experiment #3, the same variety was used with the same soil, fertilizer, pesticide and cultural management practices, but the yield was not exactly the same.

The following observations should be noted on the relationship among the 3 experiments:

- In the first experiment, the difference in yield between the two plants may be attributed to the different varieties (explained difference).
- In the second experiment, the difference in yield between the two plants may be due to the difference in soil type (explained difference).
- In the third experiment, the difference in yield between the two plants was due to reasons that are not known or cannot be explained (unexplained difference).

This is the purpose of experimental design: in conducting experiments, differences among plants (or animals or things) will occur but it is important to be able to identify what types of differences there are: one is due to the treatment applied (treatment differences), another is due to the condition in the area of the experiment (area or site differences), and last is due to conditions that are not known or cannot be explained or controlled. The first two are commonly referred to as explained variation while the last one is known as unexplained variation. In analysis of variance, we want to know how much of the total variation is due to known or explained factors (explained variance) and how much is due to unknown or unexplained factors (unexplained variance or unexplained error).

The test of significance deals with computing the ratio between the explained and unexplained variances and comparing with a probability value. If the ratio between explained and unexplained variances is relatively high, we are confident to conclude that the difference was due to the known factors and not due to unknown factors.

The purpose of proper experimental design is to make the experimental (unexplained) error small enough in relation to the treatment (explained) error. Experimental error is defined as the difference among treatments treated alike. Therefore, each treatment should be repeated several times (or replicated) in the experiment to get an estimate of this value. Replication also makes the estimate of the treatment mean more reliable.

The decision as to how many replications should there be in the experiment depends on the scientist or researcher. The factors to consider are: total available area, degree of variation within the area, amount of resources available (manpower, funding, supplies, etc.). The principle to follow is that the less the number of replications, the less precise will the conclusion be, however, too many replications may also lead to higher experimental error because the total experimental area will become larger hence more variability.

Proper randomization is necessary to prevent or minimize bias (favoring or not favoring a particular treatment over the other). With proper randomization, all treatments are given equal chance of being placed or assigned in a particular space in the experimental area (or in each block). When this is accomplished, experimental error is minimized.

The test of significance (F-test) involves comparing each of the explained variances with the corresponding unexplained variance by way of a ratio (F-computed or F_c). The bigger the F_c , the more significant are the treatment (or factor combination) differences. The F_c may become large if:

- the treatment variance is considerably large compared to the experimental error;
- the experimental error is made small in relation to the mean by proper experimental design.

Coefficient of Variation (C.V.) is the ratio between standard deviation (or square root of error mean square) and the mean. It is computed to determine whether or not to accept the results of the experiment. A very high C.V. indicates something went wrong in the conduct of the experiment: improper lay-out, too much variability in the area probably due to plot size, topography, uneven or severe insect or disease population (not part of the treatment), severe weather conditions, loss of samples due to animals, etc. When this happens, the experiment is declared invalid and the results and conclusions unacceptable. A very low C.V., on the other hand, is also dangerous because it indicates that very small differences among treatment means will be detected to be significant which may lead to a conclusion or recommendation that is not appropriate or practical.

In agricultural research, the following ranges of C.V. are deemed acceptable depending on the crop being grown:

- For field crops (rice, corn, soybean, wheat), the C.V. should not be more than 15%.
- For horticultural crops (fruit and leafy vegetables, fruit trees), the C.V. should not be more than 25%.
- For crops grown under the soil (potato, sweet potato, onion, radish, etc.), the C.V. should not be more than 40%.

Notations to be used in computation

If the rules in statistical books will be followed, the usual notation for a variable is X (or Y). Therefore, X_i refers to the value of X at the i^{th} place in the set of data. In the data on page 1, $X_1 = 4$, $X_2 = 5$, etc.

In the data for analysis of variance of single factor experiment, two-way classification is used because of the presence of replication. In this case, a variable is given the notation of X_{ij} instead of X_i alone where i refers to the i^{th} treatment and j refers to the j^{th} replication.

For two-factor experiments, three-way classification is used: one for factor A, one for factor B and one for replication. Therefore, a variable is given the notation X_{ijk} where i refers to the i^{th} level of A factor, j refers to the j^{th} the level of B factor and k refers to the k^{th} level of replication.

These notation are easy to use when someone is fully knowledgeable about statistics in general and statistical designs in particular. However, for someone who is not familiar with the topic, it may be difficult to understand, follow or even remember.

To facilitate or make understanding of statistical designs easier, the notation used by Gomez & Gomez (1984)² will be followed. In the case of single factor experiment (CRD, RCBD and Latin Square), the following notations will be used:

- $\Sigma\Sigma TR$ instead of $\Sigma\Sigma X_{ij}$ to indicate sum of all X's, in this case, the combination of treatment and replications. In the data above:

$$\Sigma\Sigma TR = T_1R_1 + T_1R_2 + T_1R_3 + T_1R_4 + \dots + T_iR_j$$

NOTE: the term $\Sigma\Sigma T_iR_j$ does not refer to the sum of the products of T and R. It refers to the value of trt 1 rep 1 plus value at trt 1 rep 2, and so on until the last value.

- $\Sigma\Sigma TR^2$ instead of $\Sigma\Sigma X_{ij}^2$ to indicate sum of the squares of individual values of T of each replication. Note that here, each value is squared first before the sum is taken.

$$\Sigma\Sigma TR^2 = (T_1R_1)^2 + (T_1R_2)^2 + (T_1R_3)^2 + (T_1R_4)^2 + \dots + (T_iR_j)^2$$

In the case of multiple factor experiments, A & B are used instead of T.

- $\Sigma\Sigma ABR$ instead of $\Sigma\Sigma X_{ijk}$ to indicate sum of all X's, in this case, the combination of Factor A, Factor B, and replications.

$$\Sigma\Sigma ABR = A_1B_1R_1 + A_1B_1R_2 + A_1B_1R_3 + A_1B_1R_4 + \dots + A_iB_jR_k$$

Rules to be followed in computing for Sum of Squares:

The most difficult part of any analysis of variance is to determine the formula of the sum of squares. An even more difficult task is to memorize or remember them. It is not advisable to memorize each and every formula simply because there are too many of them. The following technique may be used so that the proper formula of sum of squares can be derived even without remembering or memorizing them.

The first thing to study is the nature of the design in terms of number of treatments (or factors) and replications, and the relationship between or among them. It is also important to know the relative plot size of each. By knowing the nature of the design of the experiment, the proper sources of variation can be identified and the appropriate degrees of freedom established. Once the degree of freedom is set and found correct, it can be used to determine the formula for the sum of squares. The following rules should be followed.

1. Determine how many factors are involved in the experiment (treatment, replication included). Note that it is asking for the number of factors involved, not the number of levels of each factor. For simple CRD and RCBD, there will be 2 factors: treatment and replication. For Latin Square, there will be three: column, row, and treatment; for 2-factor factorial, there will be three: replication, factor A and factor B, and so on.
2. Expand all the terms in the degrees of freedom, e.g. in CRD, the expansion of the error df term $t(r-1)$ is tr - t, error df term for RCBD $(t-1)(r-1)$ is tr - t - r + 1.
3. Using the error df formula tr - t - r + 1 as an example, there are four terms: tr, -t, -r & +1.
 - a. For the first term tr, replace the small letters tr with big letters TR, square it, put double summation sign in front (one for T and one for R), and divide by the small letter(s) of the remaining factor(s). In this case, there are only two factors (t & r) and both T & R are placed above so there is no letter to be placed below therefore there is no need to divide. The formula for the first term will be:

$$\Sigma\Sigma (TR)^2 = (T_1R_1)^2 + (T_1R_2)^2 + \dots + (T_iR_j)^2$$
 - b. For the second term -t (always remember to include the sign), replace small letter t with big letter T, square it, put summation sign in front, and divide by the small letter(s) of the remaining factor(s). In this case, T is on top so r will be at the bottom. The formula of the second term will be:

² Gomez, K.A. and A.A. Gomez. 1984. Statistical Procedures for Agricultural Research. 2nd Ed. John Wiley & Sons. 680 pp.

$$- \frac{\Sigma(T)^2}{r} = - \frac{T_1^2 + T_2^2 + \dots + T_i^2}{r}$$

- c. For the third term $-r$, replace small letter r with big letter R , square it, put summation sign in front, and divide by the small letter(s) of the remaining factor(s). In this case, R is on top so t will be at the bottom. The formula of the third term will be:

$$- \frac{\Sigma(R)^2}{t} = - \frac{R_1^2 + R_2^2 + \dots + R_j^2}{t}$$

- d. The fourth term is $+1$. It refers to the Correction Factor (CF) or Correction Term (CT) which is the square of the Grand Total divided by the number of all factors (in this case, t & r), so the formula for the fourth term will be:

$$+ \frac{(GT)^2}{tr} = + \frac{(T_1R_1 + T_1R_2 + \dots + T_iR_j)^2}{t \times r}$$

where **GT** refers to the Grand Total

T_1R_1 = value of Treatment 1 in Replication 1

T_iR_j = value of the i^{th} Treatment in j^{th} Replication

The final formula for Error Sum of Squares for RCBD will be:

$$\Sigma\Sigma(TR)^2 - \frac{\Sigma(T)^2}{r} - \frac{\Sigma(R)^2}{t} + \frac{(GT)^2}{tr}$$

Note that the divisor (or denominator) will change depending on the number of factors involved in the experiment. It will be t & r for CRD and RCBD; c , r & t for Latin Square; a , b & r for two-factor experiments, a , b , c & r for three-factor experiments, and so on.

4. For more complicated error df formula such as $(a-1)(b-1)(r-1)$, there is a short-cut technique in expanding. There are 3 factors involved: **a**, **b**, and **r**. To expand:

- The first term includes all the letters with positive sign (**+abr**)
- The second group of terms includes all two letter combinations with the opposite sign from the first term (**-ab**, **-ar**, **-br**)
- The third group of terms includes all single letters with the opposite sign from the previous group (**+a**, **+b**, **+r**)
- The last term will be **1** with the opposite sign from the previous group (**-1**)

When put together, the expansion of $(a-1)(b-1)(r-1)$ will be:

$$abr - ab - ar - br + a + b + r - 1$$

If there are 4 factors (a , b , c , & r), the error df formula will be: $(a-1)(b-1)(c-1)(r-1)$

$$= abcr - abc - abr - acr - bcr + ab + ac + ar + bc + br + cr - a - b - c - r + 1.$$

Note the change in sign of the terms as the number of factors changes.

To facilitate writing of correct formula:

- Write down in one line, spaced apart, all the df including the sign. In the case of the error df formula $(a-1)(b-1)(r-1)$, the expanded formula will be:

$$abr \quad -ab \quad -ar \quad -br \quad +a \quad +b \quad +r \quad -1$$

- Convert small letter(s) into CAPITAL letter(s) together with the sign, change 1 with C.F.

$$ABR \quad -AB \quad -AR \quad -BR \quad +A \quad +B \quad +R \quad -C.F.$$

- Square each letter/group of letters (excluding C.F.)

$$(ABR)^2 \quad - (AB)^2 \quad - (AR)^2 \quad - (BR)^2 \quad + A^2 \quad + B^2 \quad + R^2 \quad - C.F.$$

- Put the appropriate number of summation signs corresponding to the number of letters

$$\Sigma\Sigma\Sigma(ABR)^2 - \Sigma\Sigma(AB)^2 - \Sigma\Sigma(AR)^2 - \Sigma\Sigma(BR)^2 + \Sigma A^2 + \Sigma B^2 + \Sigma R^2 - C.F.$$

- Divide by the remaining letter(s) of each term

$$\Sigma\Sigma\Sigma(ABR)^2 - \frac{\Sigma\Sigma(AB)^2}{r} - \frac{\Sigma\Sigma(AR)^2}{b} - \frac{\Sigma\Sigma(BR)^2}{a} + \frac{\Sigma A^2}{br} + \frac{\Sigma B^2}{ar} + \frac{\Sigma R^2}{ab} - C.F.$$

TEST OF SIGNIFICANCE OF A SINGLE POPULATION/SAMPLE

When observations are taken from the members of a population or a sample, it is possible to determine if the mean is representative of the population or the sample. Suppose a study is done to determine the average number of children per family in Kabul. The number of children of each family will be recorded, added and divided by the number of families. To determine if the average number is representative of the range of number of children, standard deviation needs to be computed where:

$$\sigma = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} = \sqrt{\frac{\sum X_i^2 - \frac{(\sum X_i)^2}{N}}{N}}$$

Even if it possible advisable to include all the families in the survey, it is not advisable because of very large number of families to be interviewed or observed which will be costly and time-consuming. Proper sampling technique is normally done and the number of children of the families in the chosen sample will be recorded, added and divided by the number of families in the sample. Standard deviation of the sample is computed using the formula:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n - 1}}$$

Consider two groups of students sampled from all students

SAMPLE	GROUP 1	GROUP 2
1	1	2
2	5	1
3	1	3
4	5	4
5	2	2
6	8	3
7	1	1
8	4	4
9	2	5
10	8	2
11	9	3
12	6	4
13	1	1
14	7	3
15	2	4
16	4	2
17	6	3
18	1	2
19	8	4
20	4	2
Mean	4.25	2.75
n	20	20

The means of the two samples are computed as:

$$\bar{X}_1 = \frac{X_{1i}}{N} = \frac{1 + 5 + \dots + 4}{20} = \frac{85}{20} = \boxed{4.25}$$

$$\bar{X}_2 = \frac{X_{2i}}{N} = \frac{2 + 1 + \dots + 2}{20} = \frac{55}{20} = \boxed{2.75}$$

The variance and standard deviation of Group 1 are computed as:

$$s_1^2 = \frac{\sum(X - \bar{X})^2}{n - 1} = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n - 1} = \frac{21^2 + 25^2 + \dots + 24^2 - \frac{(1 + 25 + \dots + 24)^2}{20}}{19} = \boxed{7.78}$$

$$s_1 = \sqrt{7.78} = \boxed{2.72}$$

Applying the same formula in Group 2:

$$s_2^2 = \boxed{1.36}$$

$$s_2 = \boxed{1.13}$$

To determine if the mean of Group 1 is a valid estimate or representative of the group, test of significance is done using the formula:

$$t_c = \frac{\bar{X}}{s} = \frac{4.25}{2.72} = \boxed{1.56}$$

The t_c is compared with the t-table value at 1% and 5% probability level and (n-1) degree of freedom which is 2.861 for 1% and 2.093 for 5%. Since the t-computed is smaller than 5% probability level in the t-table, it is concluded that the mean is not a valid estimate or representative of the group.

When the formula is applied to group 2:

$$t_c = \frac{\bar{X}}{s} = \frac{2.75}{1.13} = \boxed{2.42}$$

The t_c is compared with the t-table value at 1% and 5% probability level and (n-1) degree of freedom which is 2.861 for 1% and 2.093 for 5%. Since the t-computed value is lower than that in the t-table at the 5% probability level but higher than that at the 1% probability level, the range of values of the samples is within the mean value so the mean value is a valid estimate of the population.

TEST OF SIGNIFICANCE OF TWO POPULATIONS/SAMPLES

It is also possible to compare the means of two populations or samples to determine whether or not they are significantly different from one another. There are two ways of comparing two populations or sample:

1. Samples with equal number of observations

When testing two samples with equal number of observations, t-test is done using the formula:

$$t_c = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{s_1^2 + s_2^2}{2}}} = \frac{|4.25 - 2.75|}{\sqrt{\frac{7.78 + 1.36}{2}}} = \frac{1.50}{2.14} = \boxed{0.702}$$

The t_c is compared with the t-table value at 1% and 5% probability level and $(2n-2)$ degrees of freedom which is 2.704 for 1% and 2.021 for 5%. Since the t-computed is smaller than 5% probability level in the t-table, it is concluded that the two groups are not significantly different from each other.

2. Samples with unequal number of observations

Consider two samples with one population having 20 observations and the other 18 observations.

SAMPLE	GROUP 1	GROUP 2
1	1	2
2	5	1
3	1	3
4	5	4
5	2	2
6	8	3
7	1	1
8	4	4
9	2	5
10	8	2
11	9	3
12	6	4
13	1	1
14	7	3
15	2	4
16	4	2
17	6	3
18	1	2
19	8	
20	4	
Mean	4.25	2.72
n	20	18

When testing two samples with unequal number of observations, t-test is done using the formula:

$$t_c = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}} = \frac{|4.25 - 2.72|}{\sqrt{\frac{(20-1) \times 7.78 + (18-1) \times 2.17}{20 + 18 - 2}}} = \frac{1.53}{2.26} = \boxed{0.674}$$

The t-computed is compared with the t-table value at 5% and 1% probability level and (n_1+n_2-2) degrees of freedom. The tabular value is 2.021 at 5% and 2.704 at 1% probability levels, respectively. Since the t-computed is smaller than the t-table value at 5% probability level, it can be concluded that the two populations are not significantly different from one another.

SINGLE-FACTOR EXPERIMENTS

Experiments in which only a single factor varies while all others are kept constant are called single-factor experiments. In such experiments, the treatments consist solely of the different levels of the single variable factor. All other factors are applied uniformly to all plots at a single prescribed level. There are two groups of experimental designs that are applicable to a single-factor experiment.

1. Complete block designs. This is a group of designs which is suited for experiments with small number of treatments and is characterized by blocks, each of which contains at least one complete set of treatments.
2. Incomplete block designs. This is a group of designs which is suited for experiments with a large number of treatments and is characterized by blocks, each of which contains only a fraction of the treatments to be tested.

Because of the complexity in the computation and formula derivation, Incomplete Block designs will not be covered. The succeeding discussions will concentrate only on Complete Block designs.

Types of single-factor experiments using complete block designs

1. Completely Randomized Design (CRD)
2. Randomized Complete Block Design (RCBD)
3. Latin Square Design (LS)

COMPLETELY RANDOMIZED DESIGN (CRD)

The design is intended for experiments where there is no significant variation in the area or environment. As such, CRD is applicable for laboratory or greenhouse experiments ONLY and should not be used for field experiments. It may also be used for experiments conducted in the open provided the conditions are the same for the entire experiment. Examples are pot or seed flat experiments where the soil is thoroughly mixed. In animal experiments, the animals should be of the same age, weight, breed, etc.

The main advantage of CRD is that it can be used for experiments with equal or unequal number of treatments or vice-versa, it can be used for treatments with unequal number of replications. The main disadvantage of CRD is the limitation in its use due to the restriction of providing uniform condition in the whole experimental area.

As the name states, all experimental units (treatment and replication combinations) are completely randomized within the whole experimental area. The steps in randomization are as follows:

1. Determine the total number of experimental units. For CRD with equal number of treatments per replication, it is the product of the number of treatments and the number of replicates. For CRD with unequal number of treatments, it is the sum of the treatments of all replicates.
2. Assign a plot number to each experimental unit consecutively.
3. Assign the plot numbers to the experimental plots using a randomization scheme.
 - A. Using draw lots.
 - (1) Prepare pieces of papers corresponding to the number of experimental units.
 - (2) Write the plot number in each of the papers.
 - (3) Mix the papers thoroughly in a container.
 - (4) Without looking inside the container, draw a paper and assign the plot number on the first experimental unit. Suppose T5R1 was drawn first, it will be assigned to the first space in the experimental area.

T ₅ R ₁					

- (5) Without returning the paper previously drawn, continue drawing individual papers and assign the plot numbers until all the corresponding treatments have been assigned to all experimental unit.

When the randomization is finished, the final lay-out may look like this:

T ₅ R ₁	T ₂ R ₆	T ₄ R ₃	T ₂ R ₅	T ₃ R ₅	T ₁ R ₁
T ₂ R ₄	T ₄ R ₄	T ₅ R ₂	T ₆ R ₅	T ₅ R ₃	T ₃ R ₆
T ₆ R ₁	T ₁ R ₄	T ₁ R ₃	T ₄ R ₂	T ₆ R ₃	T ₁ R ₂
T ₃ R ₂	T ₄ R ₅	T ₆ R ₂	T ₆ R ₄	T ₆ R ₆	T ₅ R ₅
T ₁ R ₅	T ₃ R ₃	T ₅ R ₆	T ₂ R ₃	T ₅ R ₄	T ₄ R ₁
T ₃ R ₁	T ₂ R ₁	T ₄ R ₆	T ₁ R ₆	T ₃ R ₄	T ₂ R ₂

The lay-out may also be like this:

T ₁ R ₅	T ₃ R ₃	T ₅ R ₆	T ₂ R ₃	T ₅ R ₄	T ₄ R ₁	T ₂ R ₄	T ₄ R ₄	T ₅ R ₂	T ₆ R ₅	T ₅ R ₃	T ₃ R ₆
T ₅ R ₁	T ₂ R ₆	T ₄ R ₃	T ₂ R ₅	T ₃ R ₅	T ₁ R ₁	T ₆ R ₁	T ₁ R ₄	T ₁ R ₃	T ₄ R ₂	T ₆ R ₃	T ₁ R ₂
T ₃ R ₁	T ₂ R ₁	T ₄ R ₆	T ₁ R ₆	T ₃ R ₄	T ₂ R ₂	T ₃ R ₂	T ₄ R ₅	T ₆ R ₂	T ₆ R ₄	T ₆ R ₆	T ₅ R ₅

or any other form provided the conditions are basically the same in the entire experimental unit.

Note that even under greenhouse condition, shading may differ from one site to another within the same greenhouse, that is, plants placed near a wall may have less sunlight received compared to those place far from a wall. Likewise, plants placed near the window may receive more sunlight than those far from the window. Therefore, it will be better if all the treatments of an experiment are placed either near the window or away from the window whichever is suitable. It is not advisable that portion of the experiment is near the window while the other portion is far away from the window. The same principle applies to other conditions in the laboratory or greenhouse such as shading, temperature, wind direction, velocity, etc.

B. Using the table of random numbers

- (1) With eyes closed, point the finger to the table of random numbers.
- (2) Copy the middle 3 digits of all the consecutive digit numbers corresponding to the number of experimental units.
- (3) Rank the 3-digit numbers from 1 to the highest number of experimental units.
- (4) Arrange the 3-digit numbers consecutively from lowest to highest but be sure to carry the assigned rank. The sequence of the ranks will now serve as the randomized numbers.
- (5) Assign the ranks corresponding to the plot number of the experimental units.

CRD WITH EQUAL NUMBER OF REPLICATIONS

Using the following lay-out as an example, the plots may be numbered consecutively for easier data gathering:

(1) T ₅ R ₁	(2) T ₂ R ₆	(3) T ₄ R ₃	(4) T ₂ R ₅	(5) T ₃ R ₅	(6) T ₁ R ₁
(12) T ₂ R ₄	(11) T ₄ R ₄	(10) T ₅ R ₂	(9) T ₆ R ₅	(8) T ₅ R ₃	(7) T ₃ R ₆
(13) T ₆ R ₁	(14) T ₁ R ₄	(15) T ₁ R ₃	(16) T ₄ R ₂	(17) T ₆ R ₃	(18) T ₁ R ₂
(24) T ₃ R ₂	(23) T ₄ R ₅	(22) T ₆ R ₂	(21) T ₆ R ₄	(20) T ₆ R ₆	(19) T ₅ R ₅
(25) T ₁ R ₅	(26) T ₃ R ₃	(27) T ₅ R ₆	(28) T ₂ R ₃	(29) T ₅ R ₄	(30) T ₄ R ₁
(36) T ₃ R ₁	(35) T ₂ R ₁	(34) T ₄ R ₆	(33) T ₁ R ₆	(32) T ₃ R ₄	(31) T ₂ R ₂

Based on the lay-out, the basic information of the trial may be entered in the permanent record book using the format shown below. This format will facilitate easier data entry when observing the different entries.

<u>PLOT</u>	<u>TRT/REP</u>
-------------	----------------

1	T5R1
2	T2R6
3	T4R3
4	T2R5
5	T3R5
6	T1R1
7	T3R6
8	T5R3
9	T6R5
10	T5R2
11	T4R4
12	T2R4
13	T6R1
14	T1R4
15	T1R3
16	T4R2
17	T6R3
18	T1R2
19	T5R5
20	T6R6
21	T6R4
22	T6R2
23	T4R5
24	T3R2
25	T1R5
26	T3R3
27	T5R6
28	T2R3
29	T5R4
30	T4R1
31	T2R2
32	T3R4
33	T1R6
34	T4R6
35	T2R1
36	T3R1

As an example using different concentrations of NaCl on the growth medium for culturing a particular organism, the following data on the number of growing colonies were gathered:

PLOT TRT/REP NUMBER

1	T5R1	15
2	T2R6	17
3	T4R3	14
4	T2R5	15
5	T3R5	11
6	T1R1	17
7	T3R6	18
8	T5R3	12
9	T6R5	15
10	T5R2	12
11	T4R4	12
12	T2R4	11
13	T6R1	13
14	T1R4	18
15	T1R3	17
16	T4R2	22
17	T6R3	13
18	T1R2	20
19	T5R5	11
20	T6R6	16
21	T6R4	14
22	T6R2	15
23	T4R5	13
24	T3R2	22
25	T1R5	16
26	T3R3	18
27	T5R6	13
28	T2R3	19
29	T5R4	11
30	T4R1	16
31	T2R2	14
32	T3R4	14
33	T1R6	17
34	T4R6	14
35	T2R1	18
36	T3R1	18

Other sets of data may be gathered but for the example, only the number of growing colonies is observed and recorded. This procedure of using consecutive numbering will facilitate easier observation and at the same time will prevent any error in reading in case the treatment labels are lost.

For data analysis, the data has to be arranged in a simplified manner that will allow easy reading of values of each treatment. A two-way table is constructed putting together in one row all the observations for a particular treatment.

TREATMENT	REP 1	REP 2	REP 3	REP 4	REP 5	REP 6
Treatment 1	17	20	17	18	16	17
Treatment 2	18	14	19	11	15	17
Treatment 3	18	22	18	14	11	18
Treatment 4	16	22	14	12	13	14
Treatment 5	15	12	12	11	11	13
Treatment 6	13	15	13	14	15	16

After arranging all the values, the total of each treatment, total of each replication and mean of each treatment are computed as shown below.

TREATMENT	REP 1	REP 2	REP 3	REP 4	REP 5	REP 6	TOTAL	MEAN
Treatment 1	17	20	17	18	16	17	105	17.5
Treatment 2	18	14	19	11	15	17	94	15.7
Treatment 3	18	22	18	14	11	18	101	16.8
Treatment 4	16	22	14	12	13	14	91	15.2
Treatment 5	15	12	12	11	11	13	74	12.3
Treatment 6	13	15	13	14	15	16	86	14.3
TOTAL	97	105	93	80	81	95	551	15.3

Note that although there is no need to compute the Replication Totals in CRD because there is no blocking in the experiment, it is still advisable to compute for these totals to serve as double check if the Grand Total is accurate.

Degrees of Freedom (df):

$$\text{Treatment} = t - 1 = 6 - 1 = \boxed{5}$$

$$\text{Error} = t(r - 1) = 6(6 - 1) = \boxed{30}$$

$$\text{Total} = tr - 1 = 6 \times 6 - 1 = \boxed{35}$$

Applying the rules on how to compute for sum of squares discussed before, the formula for the sum of squares of each source of variation can be computed. For accuracy in the final answer, it is advisable to maintain 4 decimal places in all the answers in Sum of Squares and Mean Squares.

Correction Factor

$$\text{C.F.} = \frac{GT^2}{tr} = \frac{(551)^2}{6 \times 6} = \boxed{8,433.3611}$$

Total Sum of Squares (ToSS)

$$= \sum \sum (TR)^2 - C.F. = (T_1R_1)^2 + (T_1R_2)^2 + \dots + (T_6R_6)^2 - C.F.$$

$$= (17^2 + 20^2 + \dots + 16^2) - 8,433.3611 = 8,745.0000 - 8,433.3611 = \boxed{311.6389}$$

Treatment Sum of Squares (TrSS)

$$= \frac{\sum T^2}{r} - C.F. = \frac{T_1^2 + T_2^2 + \dots + T_6^2}{r} - C.F.$$

$$= \frac{105^2 + 94^2 + \dots + 86^2}{6} - 8,433.3611 = 8,535.8333 - 8,433.3611 = \boxed{102.4722}$$

Error Sum of Squares (ESS)

$$= \sum \sum TR^2 - \frac{\sum T^2}{r} = (T_1R_1)^2 + (T_1R_2)^2 + \dots + (T_6R_6)^2 - \frac{T_1^2 + T_2^2 + \dots + T_6^2}{r}$$

$$= (17^2 + 19^2 + \dots + 16^2) - \frac{(105^2 + 94^2 + \dots + 86^2)}{6} = 8,745.0000 - 8,535.8333 = \boxed{209.1667}$$

To double check if the computation of Sum of Squares is correct, add the Treatment SS and Error SS and compare with the Total SS. In the example

$$TrSS + ESS = 102.4722 + 209.1667 = \boxed{311.6389} \text{ so computation is correct.}$$

$$\text{Treatment Mean Square (TrMS)} = \frac{TrSS}{Tr \text{ df}} = \frac{102.4722}{5} = \boxed{20.4944}$$

$$\text{Error Mean Square (TrMS)} = \frac{ESS}{E \text{ df}} = \frac{209.1667}{30} = \boxed{6.9722}$$

$$\text{Treatment F-computed (TrFc)} = \frac{TrMS}{EMS} = \frac{20.4944}{6.9722} = \boxed{2.94}$$

After computing all the values, the Analysis of Variance (ANOVA) table can be constructed.

ANALYSIS OF VARIANCE						
Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Treatment	5	102.4722	20.4944	2.94 *	3.70	2.53
Error	30	209.1667	6.9722			
TOTAL	35	311.6389				

$$C.V. = \frac{\sqrt{EMS}}{\text{Mean}} \times 100 = \frac{\sqrt{6.9722}}{15.4} \times 100 = \boxed{17.2\%}$$

For F-computed values, it is enough to maintain two decimal places because the values in the F-table (F_t) are up to decimal places only.

To find out if the F_c is significant, highly significant or not significant, the F-table (F_t) is used for comparison. To read the F-table, the upper df refers to the Treatment df (or Block df) while the left side df refers to the Error df. In the example, the Treatment df is 5 while the Error df is 30, therefore, go to column where the df is 5 and then to the row where the df is 30. The value where these two intersect is the F_t value to compare the F_c . The value written in bold is for 1% (3.70) while the one in normal size is for 5% (2.53).

In some cases, the actual df used in the experiment is not found in the table. As an example, a CRD experiment involving 13 treatments and 4 replications will have a Treatment df of 12 ($t-1 = 13-1 = 12$) and Error df of 39 [$t(r-1) = 13 \times (4-1) = 39$]. In the table, there is no value for Error df = 39 so the average of the values of Error df 38 and 40 may be used.

To interpret results:

- If F_c is less than F_t 5%, the Treatment Means are declared not significantly different (ns). This indicates that whatever differences observed among the Treatment Means may not be attributed to the effect of the treatment alone but may also be due to unknown factors.
- If F_c is equal to or greater than F_t 5% but less than F_t 1%, the Treatment Means are declared significantly different (*). This indicates a 95% probability that the conclusion of the Treatments Means being different is correct (or 5% probability that the conclusion is wrong).
- If F_c is equal to or greater than F_t 1%, the Treatment Means are declared highly significantly different (**). This indicates a 99% probability that the conclusion of the Treatments Means being different is correct (or 1% probability that the conclusion is wrong).

When computer programs are used to do the analysis, there is no need to use the F-table because the output will show that actual probability value (Pr). Reading the values will be entirely different:

- If Pr is greater than 5%, Treatment Means differences are declared not significant.
- If Pr is less than or equal to 5% but higher than 1%, Treatment Mean differences are declared significant.
- If Pr is equal to or less than 1%, Treatment Mean differences are declared highly significant.

The results showed that there are significant differences among the Treatment Means since the F_c value is higher than the F_t at 5% but lower than the F_t at 1% probability level. The C.V. is within the acceptable limit, an indication that the conclusions drawn from the data are reliable.

It should be emphasized at this point that Analysis of Variance will detect if no significant, significant, or highly significant differences exist among the treatments being evaluated. It will NOT detect which among the treatment means are different from one another. Mean Comparison technique is needed to determine this.

Mean Comparison

There are two commonly used tests to detect mean differences among treatments: Least Significant Difference (LSD) and Duncan's Multiple Range Test (DMRT).

Least Significant Difference (LSD)

LSD is commonly used when a check or control is to be compared with the other treatments. It is limited to small number of treatments, about 4–8 only. With more than eight (8) treatments, LSD becomes less efficient. The maximum number of possible comparisons is also limited to one less the number of treatments ($t - 1$).

While it is a common practice to compare the check with the rest of the treatments when using LSD, it is possible to make different comparisons, however, the choice of the means to be

compared should be done before the experiment is to be conducted, not when the data have been gathered and analyzed. The same rule on the maximum number of possible comparison applies, that is, it should not exceed the treatment degrees of freedom ($t - 1$).

Formula for LSD:

$$\text{LSD} = t_{(\text{Edf}, 0.05)} \times s_d$$

where t = tabular value at 5% probability level at Error df

$$s_d = \text{standard error of mean difference} = \sqrt{\frac{2\text{MSE}}{r}}$$

Using the results of the Complete Randomized Design where Error Mean Square is 6.9722, and using the 5% t -table value for error df of 30 = 2.032 and 1% = 2.727:

$$s_d = \sqrt{\frac{2 \times 6.9722}{6}} = \boxed{1.5245}$$

$$\text{LSD}_{0.05} = s_d \times t_{0.05, 30 \text{ df}} = 1.5245 \times 2.032 = \boxed{3.1}$$

$$\text{LSD}_{0.01} = s_d \times t_{0.01, 30 \text{ df}} = 1.5245 \times 2.727 = \boxed{4.2}$$

Depending on the desired confidence level, the treatment means may be compared using the computed LSD values. Confidence level refers to $(100 - \alpha)$ assurance that the conclusion made is correct:

- At $\alpha = 0.05$, the researcher is 95% confident that his conclusion is true based on the statistical analysis.
- At $\alpha = 0.01$, the researcher is 99% confident that his conclusion is true based on the statistical analysis.

In agriculture, up to 95% confidence level is accepted. Anything below 95% confidence level is deemed not safe to make a valid conclusion. In other disciplines (social science, biological science), confidence level may be different (up to 90% sometimes is acceptable).

Using the results above and assuming that Treatment 6 is the check or control, means of Treatments 1 to 5 will be compared to the mean of Treatment 6. The treatment means and the absolute difference from the check are as follows:

TREATMENT	MEAN	ABSOLUTE DIFF. FROM CHECK
Treatment 1	17.5 *	3.2
Treatment 2	15.7 ns	1.4
Treatment 3	16.8 ns	2.5
Treatment 4	15.2 ns	0.9
Treatment 5	12.3 ns	2.0
Treatment 6 (chk)	14.3	

Compare the absolute difference with the LSD value. Since the ANOVA declared that mean differences are significant at 5% but not at 1% level, LSD at 5% level is chosen. Any mean difference greater than the computed LSD value indicates significant difference from the check or control.

In this example, only Treatment 1 mean is significantly different from the check since the difference is bigger than the LSD value ($3.2 > 3.1$). All other treatments are not significantly different from the check.

Duncan's Multiple Range Test (DMRT)

For experiments that require the comparison of all possible pairs of treatment means, the LSD test is suitable because of its restriction on the maximum number of pair tests that can be done ($= t - 1$). In such a case, Duncan's multiple range test (DMRT) is the appropriate test to use.

The procedure for applying the DMRT is similar to that of LSD test; DMRT involves the computation of numerical boundaries for the classification of the difference between any two treatment means as significant or non-significant. However, unlike the LSD test in which only a single value is required for any pair comparison at a prescribed level of significance, the DMRT requires computation of a series of values, each corresponding to a specific set of pair of comparisons.

The procedure for computing the DMRT values, as for the LSD test, depends primarily on the specific ***S_d*** of the pair of treatments being compared. For our example, the six treatment means are:

TREATMENT	MEAN
Treatment 1	17.5
Treatment 2	15.7
Treatment 3	16.8
Treatment 4	15.2
Treatment 5	12.3
Treatment 6 (chk)	14.3

The steps to follow are:

1. Rank all the treatment means in decreasing (or increasing) order. It is customary to rank the treatment means according to the order of importance. For yield data, means are usually ranked from the highest-yielding treatment to the lowest-yielding treatment. For data on pest incidence, means are usually ranked from the least-infested treatment to the most severely infested treatment. The ranking based on highest to lowest will be:

Treatment 1	17.5
Treatment 3	16.8
Treatment 2	15.7
Treatment 4	15.2
Treatment 6 (chk)	14.3
Treatment 5	12.3

2. Compute the *S_d* value following the appropriate procedures for specific designs. For the example, *S_d* is computed as:

$$s_d = \sqrt{\frac{2 \times 6.9722}{6}} = \boxed{1.5245}$$

3. Compute the (*t*–1) values of the *shortest significant ranges* as:

$$Rp = \frac{(r_p)(s_d)}{\sqrt{2}} \quad \text{for } p = 2, 3, \dots, t$$

where:

t is the total number of treatments

S_d is the standard error of the mean difference computed in step 2

r_p values are the tabular values of the *significant studentized ranges* obtained from the table on Significant Studentized Ranges for New Multiple Range Test.

p is the difference in rank between the pairs of treatment means to be compared (i.e. $p = 2$ for the two means with consecutive rankings and $p = t$ for the highest and lowest means).

4. For our example, the r_p values with error df of 30 and at the 5% level of significance are:

p	$r_p(0.05)$
2	2.89
3	3.04
4	3.12
5	3.20
6	3.25

The corresponding Rp values are:

$$Rp = \frac{(r_p)(s_d)}{\sqrt{2}}$$

$$2 = \frac{2.89 \times 1.5245}{\sqrt{2}} = \boxed{3.1}$$

$$3 = \frac{3.04 \times 1.5245}{\sqrt{2}} = \boxed{3.3}$$

$$4 = \frac{3.12 \times 1.5245}{\sqrt{2}} = \boxed{3.3}$$

$$5 = \frac{3.20 \times 1.5245}{\sqrt{2}} = \boxed{3.4}$$

$$6 = \frac{3.25 \times 1.5245}{\sqrt{2}} = \boxed{3.5}$$

5. To determine whether two means are significantly different from one another, the respective Rp values should be used. In the example, the Treatment Means in descending order are:

Treatment	Mean	Difference	Rp value
Treatment 1	17.5		
Treatment 3	16.8	0.7	3.1
Treatment 2	15.7	1.8	3.3
Treatment 4	15.2	2.3	3.3
Treatment 6 (chk)	14.3	3.2	3.4
Treatment 5	12.3	5.2	3.5

- Compare the difference between Trt 1 with Trt 3 using Rp 2 value (3.1). Since the difference between the two treatments is 0.7 (< 2.2), then Trt 3 is declared not significantly different from Treatment 1.
- Compare the difference between Trt 1 and Trt 2 with Rp 3 value (3.3). Since $1.8 < 2.3$, Trt 2 is declared not significantly different from Trt 1.
- Compare the difference between Trt 1 and Trt 4 with Rp 4 value (3.3). Since $2.3 < 3.3$, Trt 4 is declared not significantly different from Trt 1.
- Compare the difference between Trt 1 and Trt 6 with Rp 5 value (3.4). Since $3.2 < 3.3$, Trt 6 is declared not significantly different from Trt 1.

- Compare the difference between Trt 1 and Trt 5 with Rp 6 value (3.5). Since $5.2 > 3.3$, Trt 5 is declared significantly different from Trt 1.
- When a significant difference is detected, there is no need to compare the rest of the treatments below as all of them will be significantly different from the treatment being compared.
- Vertical line notation is used to identify which treatment means are not significantly different:

Treatment 1	17.5	
Treatment 3	16.8	
Treatment 2	15.7	
Treatment 4	15.2	
Treatment 6 (chk)	14.3	
Treatment 5	12.3	

The first line indicates that Trt 1 is not significantly different from Trts 3, 2, 4 & 6 but significantly different from Trt 5.

- Compare Treatment 3 with the rest of the Treatment means below it.

Treatment	Mean	Difference	Rp value
Treatment 3	16.8		
Treatment 2	15.7	1.2	3.1
Treatment 4	15.2	1.7	3.3
Treatment 6 (chk)	14.3	2.5	3.3
Treatment 5	12.3	4.5	3.4

- Compare the difference between Trt 3 and Trt 2 with Rp 2 value (3.1). Since $1.2 < 2.2$, Trt 2 is declared not significantly different from Trt 3.
- Compare the difference between Trt 3 and Trt 4 with Rp 3 value (3.3). Since $1.7 < 2.3$, Trt 3 is declared not significantly different from Trt 4.
- Compare the difference between Trt 3 and Trt 6 with Rp 4 value (3.3). Since $2.5 < 2.3$, Trt 3 is declared not significantly different from Trt 6.
- Compare the difference between Trt 3 and Trt 6 with Rp 5 value (3.4). Since $4.5 > 2.3$, Trt 3 is declared significantly different from Trt 5.

Treatment 1	17.5	
Treatment 3	16.8	
Treatment 2	15.7	
Treatment 4	15.2	
Treatment 6 (chk)	14.3	
Treatment 5	12.7	

The second line indicates that Trt 3 is not significantly different from Trts 2 & 4 but significantly different from Trts 6 & 5

- Continue with the other comparisons until Trt 6 is compared to Trt 5

7. When completed, the final line notation will be:

Treatment 1	17.5			
Treatment 3	16.8			
Treatment 2	15.7			
Treatment 4	15.2			
Treatment 6 (chk)	14.3			
Treatment 5	12.3			

8. Replace the vertical lines with letters starting from letter **a**. Means that are not significantly different should be assigned the same letter. In the first comparison, Trt 1 is not significantly different from Trts 3, 2, & 4 but significantly different from Trt 6 & 5 so Trts 1, 3, 2, & 4 are assigned a common letter **a**. Initially:

Treatment 1	17.5 a
Treatment 3	16.8 a
Treatment 2	15.7 a
Treatment 4	15.2 a
Treatment 6 (chk)	14.3 a
Treatment 5	12.3

9. Trts 3, 2, & 4 are not significantly different so they are assigned a common letter **b**.

Treatment 1	17.5 a
Treatment 3	16.8 ab
Treatment 2	15.7 ab
Treatment 4	15.2 ab
Treatment 6 (chk)	14.3 ab
Treatment 5	12.3

10. Continue with the letter assignment for the other comparisons.

Treatment 1	17.5 a
Treatment 3	16.8 ab
Treatment 2	15.7 abc
Treatment 4	15.2 abcd
Treatment 6 (chk)	14.3 abcd
Treatment 5	12.3 d

11. Since Treatments 1, 3, 2, 4 & 5 are not significantly different, the 3 letters (a, b & c) assigned to them can be reduced to just 1 letter (**a**). The fourth vertical line initially assigned with letter **d** will be assigned letter **b**. The final line notation will be:

Treatment 1	17.5 a
Treatment 3	16.8 a
Treatment 2	15.7 a
Treatment 4	15.2 ab
Treatment 6 (chk)	14.3 ab
Treatment 5	12.3 b

CRD with Unequal Number of Replications

The reason for having unequal number of replications may be due to limitation in the number of available sample plants at the start of the experiment. It is also possible that equal number of plants was used at the start of the experiment but some of the plants died due to reasons aside from the treatment itself resulting to unequal number of treatments. NOTE that if one or some of the plants died due to the effect of the treatment, e.g. high concentration of a particular chemical, and not due to other unknown factors, the data on the dead plant should be recorded as zero and the analysis should be done using equal number of treatments.

A possible lay-out for CRD with unequal number of treatments is as follows:

T ₁ R ₁	T ₂ R ₅	T ₄ R ₃	T ₂ R ₂	T ₃ R ₁	T ₄ R ₂
T ₂ R ₆	T ₄ R ₄	T ₅ R ₁	T ₆ R ₂	T ₅ R ₃	T ₃ R ₃
T ₆ R ₁	T ₁ R ₂	T ₂ R ₃	T ₄ R ₆	T ₁ R ₄	T ₅ R ₄
T ₃ R ₃	T ₄ R ₅	T ₆ R ₃	T ₆ R ₅	T ₆ R ₄	
T ₅ R ₂	T ₃ R ₄	T ₅ R ₅	T ₂ R ₁	T ₅ R ₆	
T ₃ R ₅	T ₂ R ₄	T ₄ R ₁	T ₁ R ₃	T ₃ R ₆	

In the hypothetical lay-out, Treatment 1 has only 4 replications, Treatment 6 has 5 replications while the other 4 treatments have 6 replications. Again, the lay-out may be arranged in any manner (shape or orientation) provided the conditions for all the plants are basically the same or equal. A set of hypothetical data involving six treatments with 4 to 6 replications will be used.

TREATMENT	REP 1	REP 2	REP 3	REP 4	REP 5	REP 6	TOTAL	MEAN
Treatment 1	17	20	17	18			72	18.0
Treatment 2	18	14	19	11	15	17	94	15.7
Treatment 3	18	22	18	14	11	18	101	16.8
Treatment 4	16	22	14	12	13	14	91	15.2
Treatment 5	15	12	12	11	11	13	76	12.3
Treatment 6	13	15	13	14	15		70	14.0
TOTAL	97	105	93	80	65	62	502	15.2

Again, even if the Replication Totals are not needed in the computation, it is advisable to compute for them to double check if the Grand Total is accurate. Note also that the treatment means are averages based on individual number of replications while the grand mean is the overall average of the 33 samples. DO NOT compute the average of the 6 Treatment Means.

Degrees of Freedom

$$\begin{aligned}
 \text{Treatment} &= t - 1 = 6 - 1 = 5 \\
 \text{Error} &= \sum(t_r - 1) = (4 - 1) + (6 - 1) + \dots + (5 - 1) = 27 \\
 \text{Total} &= \sum t_r - 1 = (4 + 6 + 6 + 6 + 6 + 5) - 1 = 32
 \end{aligned}$$

Correction Factor (C.F.)

$$= \frac{(GT)^2}{(\sum t)} = \frac{(502)^2}{33} = 7,636.4848$$

Sums of Squares

Total (ToSS)

$$= \sum \sum (TR)^2 - C.F. = (T_1R_1)^2 + (T_1R_2)^2 + \dots + (T_6R_6)^2 - C.F.$$

$$= 17^2 + 20^2 + \dots + 15^2 - 7,636.4848 = 7,944.0000 - 7,636.4848 = \boxed{307.5152}$$

Treatment (TrSS)

$$= \sum \frac{T^2}{r_{ti}} - C.F. = \frac{T_1^2}{r_{t1}} + \frac{T_2^2}{r_{t2}} + \dots + \frac{T_6^2}{r_{t6}} - C.F.$$

$$= \frac{72^2}{4} + \frac{94^2}{6} + \dots + \frac{70^2}{5} - 7,636.4848 = 7,741.6667 - 7,636.4848 = \boxed{105.1818}$$

Error (ESS)

$$= \sum \sum TR^2 - \sum \frac{T^2}{r_t} = (T_1R_1)^2 + (T_1R_2)^2 + \dots + (T_6R_6)^2 - \frac{T_1^2}{r_{t1}} + \frac{T_2^2}{r_{t2}} + \dots + \frac{T_6^2}{r_{t6}}$$

$$= 17^2 + 20^2 + \dots + 13^2 - \frac{72^2}{4} + \frac{94^2}{6} + \dots + \frac{70^2}{5} = 7,944.0000 - 7,741.6667 = \boxed{202.3333}$$

Mean Squares

$$\text{Treatment Mean Square (TrMS)} = \frac{\text{TrSS}}{\text{Tr df}} = \frac{105.1818}{5} = \boxed{21.0364}$$

$$\text{Error Mean Square (EMS)} = \frac{\text{ESS}}{\text{E df}} = \frac{202.3333}{27} = \boxed{7.4938}$$

F-computed

$$\text{Treatment F-computed (Tr Fc)} = \frac{\text{TrMS}}{\text{EMS}} = \frac{21.0364}{7.4938} = \boxed{2.81}$$

After computing all the values, the Analysis of Variance Table can be constructed.

ANALYSIS OF VARIANCE						
Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Treatment	5	105.1818	21.0364	2.81 *	3.79	2.57
Error	27	202.3333	7.4938			
TOTAL	32	307.5152				

$$C.V. = \frac{\sqrt{\text{EMS}}}{\text{Mean}} \times 100 = \frac{\sqrt{7.4938}}{15.2} \times 100 = \boxed{18.0\%}$$

The C.V. value is within the acceptable limit so the experiment results are valid. Fc value is higher than Ft 5% so the Treatment Means are declared significantly different. This indicates that differences observed among the Treatment means can be attributed to the effect of the treatments alone and are not significantly affected by unknown or uncontrolled factors.

Mean Comparison

Because of different number of observations per treatment, LSD is recommended to test significance of difference between two treatment means.

$$\text{LSD} = t_{(\text{Edf}, 0.05)} \times s_d$$

where t = tabular value at 5% probability level at error df

$$s_d = \text{standard error of mean difference} = \sqrt{s^2 \left[\frac{1}{r_i} + \frac{1}{r_j} \right]}$$

where r_i and r_j are the replication numbers of the treatment means being compared

In the example, Treatment 6 (check with 5 replications) will be compared with the rest of the treatments. The LSD value to compare Treatment 6 with Treatment 1 (with 4 replications) will be different from the LSD value for the other treatments.

Treatment 6 vs. Treatment 1:

$$S_d = \sqrt{7.4938 \left[\frac{1}{5} + \frac{1}{4} \right]} = \boxed{1.836}$$

$$\text{LSD } 5\% = 1.836 \times 2.062 = \boxed{3.8}$$

The absolute difference between Treatment 1 and Treatment 6 means is 4.0 which is higher than LSD value of 3.8. Therefore, the 2 treatments are declared significantly different.

Treatment 6 vs. other treatments:

$$S_d = \sqrt{7.4938 \left[\frac{1}{5} + \frac{1}{6} \right]} = \boxed{1.657}$$

$$\text{LSD } 5\% = 1.657 \times 2.062 = \boxed{3.4}$$

The range of difference between Treatment 6 and the other treatments is 1.2 to 2.8 which is smaller than the LSD value so means of Treatments 2, 3, 4 and 5 are declared not significantly different from mean of Treatment 6.

RANDOMIZED COMPLETE BLOCK DESIGN

Experiments in the open field is conducted using Randomized Complete Block Design (RCBD) because it is difficult to totally control the condition in the experimental area. Variation in the condition of the area may be due to the soil itself (fertility, soil type), slope or gradient, wind direction, water direction, presence (or absence) of disease and insect pests, etc. Because of these uncontrolled factors, Completely Randomized Design is not applicable. With RCBD, blocking is introduced which will help reduce (not totally eliminate) the effect of the uncontrolled factors.

RCBD is considered to be powerful among the basic designs simply because it is able to partition the total variance into the effect of the treatment, the effect of the block and the unexplained error. Blocking is a method of improving accuracy of an experiment by arranging the experimental materials into groups so that the units in each group are as homogeneous (uniform) as possible and the treatments are arranged so that the treatment comparisons are made within each group, thereby eliminating the variability between groups.

The basic difference between CRD and RCBD is the presence of blocks in RCBD. In CRD, randomization is done in the whole experimental area while in RCBD, randomization is done in each block and each randomization is independent of the other. Since randomization is done in each block, all treatments should appear in each block (or each treatment should appear only once in each block).

In randomization and layout, two important decisions have to be made in arriving at an appropriate and effective blocking technique:

- The selection of the source of variability to be used as the basis for blocking.
- The selection of block shape and orientation.

An ideal source of variation to use as the basis for blocking is one that is large and highly predictable. Examples are:

- Soil heterogeneity or variability, in a fertilizer or variety trial where yield data is the primary character of interest.
- Direction of insect migration, in an insecticide trial where insect infestation is the primary character of interest.
- Slope of the field, in a study of plant reaction to water stress.

With these things in mind, blocking can now be done so that the variability of the area of a particular block is made as small as possible. The guidelines for blocking are as follows:

- When the gradient is unidirectional, use long and narrow blocks so that their length is perpendicular to the gradient.
- When the fertility gradient occurs in two directions with one gradient much stronger than the other, ignore the weaker gradient and follow the preceding guideline for the case of unidirectional gradient.
- When the fertility gradient occurs in two directions with both gradients equally strong and perpendicular to each other, choose one of the alternatives:
 - Use blocks that are as square as possible.
 - Use long and narrow blocks with their lengths perpendicular to the direction of one gradient and use covariance technique to take care of the other gradient.
 - Use the Latin Square Design with two-way blockings, one for each gradient.

Randomization

The steps in randomization for RCBD are as follows:

1. Divide the experimental area into the desired number of blocks. In the example, there are 6 blocks with 6 treatments per block. The shape and orientation of blocks will depend on the factors that will reduce or minimize intrablock error while maximizing interblock error.

Suppose the blocks will be arranged like this:

BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4	BLOCK 5	BLOCK 6

2. Assign a plot number to each treatment consecutively in each block.
3. Assign the treatments to the experimental plots using a randomization scheme.
 - A. Using the table of random numbers
 - (1) With eyes closed, point the finger to the table of random numbers.
 - (2) Copy the middle 3 digits of all the consecutive 5--digit numbers corresponding to the number of treatments in a block.
 - (3) Rank the 3-digit numbers from 1 to the highest number of treatments.
 - (4) Arrange the 3-digit numbers consecutively from lowest to highest but be sure to carry the assigned rank. The sequence of the ranks will now serve as the randomized numbers.
 - (5) Assign the ranks corresponding to the number of treatments of the first block.

Trt 1					
Trt 3					
Trt 4					
Trt 2					
Trt 6					
Trt 5					
BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4	BLOCK 5	BLOCK 6

- (6) Repeat steps 1-5 to the remaining blocks. Once finished, the lay-out may look like this:

Trt 1	Trt 2	Trt 6	Trt 3	Trt 5	Trt 2
Trt 3	Trt 4	Trt 3	Trt 1	Trt 4	Trt 1
Trt 4	Trt 6	Trt 1	Trt 4	Trt 1	Trt 4
Trt 2	Trt 1	Trt 2	Trt 5	Trt 2	Trt 3
Trt 6	Trt 3	Trt 5	Trt 6	Trt 3	Trt 6
Trt 5	Trt 5	Trt 4	Trt 2	Trt 6	Trt 5
BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4	BLOCK 5	BLOCK 6

If the fertility of the area is not known, the blocks and plots may be arranged this way:

BLOCK 5	Trt 1	Trt 2	Trt 3	Trt 3	Trt 1	Trt 5	BLOCK 6
	Trt 6	Trt 4	Trt 5	Trt 6	Trt 2	Trt 4	
BLOCK 3	Trt 2	Trt 1	Trt 6	Trt 4	Trt 5	Trt 2	BLOCK 4
	Trt 3	Trt 5	Trt 4	Trt 1	Trt 6	Trt 4	
BLOCK 1	Trt 5	Trt 4	Trt 2	Trt 6	Trt 3	Trt 5	BLOCK 2
	Trt 1	Trt 6	Trt 3	Trt 6	Trt 2	Trt 1	

B. Using draw lots

- (1) Prepare pieces of papers corresponding to the number of treatments in each block.
- (2) Write the treatment name in each of the paper.
- (3) Mix the papers thoroughly in a container.
- (4) Without looking in the container, draw a paper and assign the treatment name on the first experimental unit of the first block.
- (5) Without returning the paper previously drawn, continue drawing individual papers and assign the corresponding treatments until all the treatments have been assigned to the first block.
- (6) Repeat steps 1-5 to the remaining blocks. Do the same for the remaining blocks.

Analysis of Variance

The hypothetical data used in CRD will also be used in RCBD to have a comparison between the two designs. The data are as follows:

TREATMENT	REP 1	REP 2	REP 3	REP 4	REP 5	REP 6	TOTAL	MEAN
Treatment 1	17	20	17	18	16	17	105	17.5
Treatment 2	18	14	19	11	15	17	94	15.7
Treatment 3	18	22	18	14	11	18	101	16.8
Treatment 4	16	22	14	12	13	14	91	15.2
Treatment 5	15	12	12	11	11	13	74	12.3
Treatment 6	13	15	13	14	15	16	86	14.3
TOTAL	97	105	93	80	81	95	551	15.3

Degrees of Freedom (df):

Treatment	= $t - 1$	= $6 - 1 = 5$
Block	= $(r - 1)$	= $6 - 1 = 5$
Error	= $(t - 1)(r - 1)$	= $(6 - 1)(6 - 1) = 25$
Total	= $tr - 1$	= $6 \times 6 - 1 = 35$

Sum of Squares

Correction Factor

$$C.F. = \frac{GT^2}{tr} = \frac{(551)^2}{6 \times 6} = \boxed{8,433.3611}$$

Total Sum of Squares (ToSS)

$$= \sum \sum (TR)^2 - C.F. = (T_1R_1)^2 + (T_1R_2)^2 + \dots + (T_6R_6)^2 - C.F.$$

$$= (17^2 + 20^2 + \dots + 15^2) - 8,433.3611 = 8,745.0000 - 8,433.3611 = \boxed{311.6389}$$

Block (Replication) Sum of Squares (RSS)

$$= \frac{\sum R^2}{t} - C.F. = \frac{R_1^2 + R_2^2 + \dots + R_6^2}{t} - C.F.$$

$$= \frac{97^2 + 105^2 + \dots + 95^2}{6} - 8,433.3611 = 8,511.5000 - 8,433.3611 = \boxed{78.1389}$$

Treatment Sum of Squares (TrSS)

$$= \frac{\sum T^2}{r} - C.F. = \frac{T_1^2 + T_2^2 + \dots + T_6^2}{r} - C.F.$$

$$= \frac{105^2 + 94^2 + \dots + 86^2}{6} - 8,433.3611 = 8,535.8333 - 8,433.3611 = \boxed{102.4722}$$

Error Sum of Squares (ESS)

$$= \sum \sum TR^2 - \frac{\sum T^2}{r} - \frac{\sum R^2}{t} + C.F.$$

$$= (17^2 + 20^2 + \dots + 15^2) - \frac{(105^2 + 94^2 + \dots + 85^2)}{6} - \frac{(97^2 + 105^2 + \dots + 94^2)}{6} + 8,433.3611$$

$$= 8,745.0000 - 8,535.8333 - 8,511.5000 + 8,433.3611 = \boxed{131.0278}$$

Mean Squares

Treatment Mean Square (TrMS) = $\frac{TrSS}{Tr \text{ df}} = \frac{102.4722}{5} = \boxed{20.4944}$

Block Mean Square (RMS) = $\frac{RSS}{R \text{ df}} = \frac{78.1389}{5} = \boxed{15.6278}$

$$\text{Error Mean Square (EMS)} = \frac{\text{ESS}}{\text{E df}} = \frac{131.0278}{25} = \boxed{5.2411}$$

F-computed

$$\text{Block F-computed (R Fc)} = \frac{\text{RMS}}{\text{EMS}} = \frac{15.6278}{5.2411} = \boxed{2.98}$$

$$\text{Treatment F-computed (Tr Fc)} = \frac{\text{TrMS}}{\text{EMS}} = \frac{20.4944}{5.2411} = \boxed{3.91}$$

After computing all the values, the Analysis of Variance Table can be constructed.

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	5	78.1389	15.6278	2.98 *	3.86	2.60
Treatment	5	102.4722	20.4944	3.91 **	3.86	2.60
Error	25	131.0278	5.2411			
TOTAL	35	311.6389				

$$\text{C.V.} = \frac{\sqrt{(\text{EMS})}}{\text{Mean}} \times 100 = \frac{\sqrt{5.2411}}{15.3} \times 100 = \boxed{15.0\%}$$

By analyzing the same set of data using RCBD, the treatment and block mean differences are significant. Because of the addition of block as source of variation, the residual error mean square was reduced from 6.9722 in CRD to 5.2411. This indicates that blocking was effective and block orientation was proper since it managed to absorb the bulk of the unexplained error resulting to lower EMS and consequently, higher Fc. Because of lower EMS, a lower C.V. of 15.0% was observed in RCBD compared to 17.2% in CRD. This is also reflected in Relative Efficiency of RCBD over CRD using the formula:

$$\text{R.E.} = \frac{(r-1)E_b + r(t-1)E_e}{(tr-1)E_e} = \frac{(6-1) \times 15.6278 + 6(6-1) \times 5.2411}{(6 \times 6 - 1) \times 5.2411} = \boxed{1.28}$$

By using RCBD, the experimental precision was increased by 28% compared to when using CRD.

Mean Comparison

1. Least Significant Difference

The same LSD formula used in CRD are applicable to RCBD when comparing treatment means. The *t* value will change though because of the reduction in error df.

$$s_d = \frac{\sqrt{2 \times 5.2411}}{6} = \boxed{1.3218}$$

$$\text{LSD}_{0.05} = s_d \times t_{0.05, 25 \text{ df}} = 1.5172 \times 2.060 = \boxed{2.7}$$

Since ANOVA declared that the Treatment means are significantly different, LSD 5% will be used to compare Treatment means. With 6 treatments, 5 valid pair comparisons can be made. Comparing the check (Treatment 6) with the other treatments, the results showed the following:

Treatment	Mean	Difference
Treatment 1	17.5 *	3.3
Treatment 2	15.7 ns	1.5
Treatment 3	16.8 ns	2.7
Treatment 4	15.2 ns	1.0
Treatment 5	12.7 ns	1.5
Treatment 6 (check)	14.2	

Applying LSD 5%, only Treatment 1 is significantly different from the check.

2. DMRT

The same procedure in computation of DMRT as in CRD is followed but the Rp values will change because of the change in Error df.

When completed, the final DMRT table will be:

Treatment 1	17.5 a
Treatment 3	16.8 ab
Treatment 2	15.7 abc
Treatment 4	15.2 abc
Treatment 6 (chk)	14.3 cd
Treatment 5	12.3 d

To discuss and conclude on the results, Treatments 1 & 3 are significantly different from Treatment 6 & 5 since they do not have any common letter; Treatment 2 & 4 are significantly different from Treatment 6 for the same reason; Treatments 6 & 5 are not significantly different.

LATIN SQUARE DESIGN

This design is most useful in areas where the direction of soil fertility/heterogeneity is bi-directional. With this type of soil fertility, it is not advisable to use RCBD because the blocking will take care of only one gradient while the other gradient will be confounded (or added) to the treatment effect. As a result, the variation observed among treatments cannot be attributed to the effect of the treatment alone because part of the variation may be due to differences in fertility gradient. Latin Square is the more appropriate design because the two-directional blocking, commonly referred to as row-blocking and column-blocking, is accomplished by ensuring that every treatment occurs only once in each row-block and once in each column-block. As such, LS design is considered more powerful than RCBD in that aside from detecting differences due to treatments, it also detects differences due to rows and columns and not due to blocks alone.

Some examples of cases where LS design can be appropriately used are:

- Field trials in which the experimental area has two fertility gradient running perpendicular to each other, or has a unidirectional fertility gradient but also has a residual effects from previous trials.
- Insecticide field trials where the insect migration has a predictable migration which is perpendicular to the dominant fertility gradient of the experimental area.
- Greenhouse trials where the experimental pots are arranged in straight line which is perpendicular to the glass or screen walls, such that the differences among rows of pots and the distance from the glass wall (or screen wall) are expected to be the two major sources of variability among the experimental pots.
- Laboratory trials with replication over time, such that the difference among experimental units conducted at the same time and among those conducted over time constitute the two known sources of variability.

One important feature of the design is that the number of replications is always equal to the number of treatments. As such, the LS design can only be used when the number of treatments is a perfect square (9, 16, 25, 36, 49, etc.). Because of this requirement, the main disadvantage of the design is that it is not advisable for large number of treatments. Thus, in practice, the LS design is applicable only for experiments in which the number of treatments is not less than four but not more than eight. Because of such limitation, the LS design has not been widely used in agricultural experiments despite its great potential for controlling experimental error.

Randomization and lay-out

The process of randomization and lay-out for a LS design is shown below for an experiment with six treatments, A, B, C, D, E and F.

1. Consider a sample LS plan with six treatments. For our example, the 6 x 6 Latin Square preliminary plan is:

A	B	C	D	E	F
F	A	B	C	D	E
E	F	A	B	C	D
D	E	F	A	B	C
C	D	E	F	A	B
B	C	D	E	F	A

2. Randomize the row arrangement of the plan selected in step 1, following one of the randomization schemes described in RCBD. For this experiment, the table-of-random numbers method is applied.
 - With eyes closed, point the finger to the table of random numbers.
 - Copy the middle 3 digits of all five consecutive 5-digit numbers corresponding to the number of treatments. For our example, 717, 569, 223, 478, 036 and 132 were taken.

- Rank the 3-digit numbers from lowest to highest:

Random Number	Sequence	Rank
717	1	6
569	2	5
223	3	3
478	4	4
036	5	1
132	6	2

- Use the rank to represent the existing row number of the selected plan and the sequence to represent the row number of the new plan. For our example, the fifth row of the selected plan (rank = 1) becomes the first row (sequence = 1) of the new plan; the sixth row of the selected plan becomes the second row of the new plan, and so on. The new row plan, after the row randomization, is:

C	D	E	F	A	B
B	C	D	E	F	A
E	F	A	B	C	D
D	E	F	A	B	C
F	A	B	C	D	E
A	B	C	D	E	F

- Randomize the column arrangement, using the same procedure used for row arrangement in step 2. For our example, the six random numbers selected and their ranks are:

Random Number	Sequence	Rank
092	1	1
817	2	3
978	3	6
860	4	5
848	5	4
709	6	2

The rank will now be used to represent the column number of the plan obtained in step 2 (i.e. with re-arranged rows) and the sequence will be used to represent the column number of the new plan.

For our example, the first column obtained in step 2 remains the first column of the final plan, the sixth column of the plan of step 2 becomes the second column of the final plan, and so on. The final plan, which becomes the lay-out of the experiment is:

Column						
Row	1	2	3	4	5	6
1	C	B	D	A	F	E
2	B	A	C	F	E	D
3	E	D	F	C	B	A
4	D	C	E	B	A	F
5	F	E	A	D	C	B
6	A	F	B	E	D	C

Consequently, the field lay-out is shown below. Note that the dimension of the lay-out cannot be changed unlike in CRD or RCBD.

	COL 1	COL 2	COL 3	COL 4	COL 5	COL 6
ROW 1	Trt 3	Trt 2	Trt 4	Trt 1	Trt 6	Trt 5
ROW 2	Trt 2	Trt 1	Trt 3	Trt 6	Trt 5	Trt 4
ROW 3	Trt 5	Trt 4	Trt 6	Trt 3	Trt 2	Trt 1
ROW 4	Trt 4	Trt 3	Trt 5	Trt 2	Trt 1	Trt 6
ROW 5	Trt 6	Trt 5	Trt 1	Trt 4	Trt 3	Trt 2
ROW 6	Trt 1	Trt 6	Trt 2	Trt 5	Trt 4	Trt 3

Analysis of Variance

The same set of hypothetical data used in CRD and RCBD involving 6 treatments (designated by letters in parenthesis) will be used with the assigned columns and rows included:

ROW	COLUMN						Row Total	Trt Total
	1	2	3	4	5	6		
1	(A) 17	(F) 15	(C) 18	(D) 12	(E) 11	(B) 17	90	(A) 105
2	(B) 18	(C) 22	(E) 12	(F) 14	(A) 16	(D) 14	96	(B) 94
3	(C) 18	(D) 22	(B) 19	(A) 18	(F) 15	(E) 13	105	(C) 101
4	(D) 16	(A) 20	(F) 13	(E) 11	(B) 15	(C) 18	93	(D) 91
5	(E) 15	(B) 14	(A) 17	(C) 14	(D) 13	(F) 16	89	(E) 75
6	(F) 13	(E) 12	(D) 14	(B) 11	(C) 11	(A) 17	78	(F) 86
Column Total	97	105	93	80	81	94		551

Degrees of Freedom (df):

Treatment	= $t - 1$	= $6 - 1$	= 5
Column	= $c - 1$	= $6 - 1$	= 5
Row	= $r - 1$	= $6 - 1$	= 5
Error	= $(t - 1)(t - 2)$	= $(6 - 1)(6 - 2)$	= 20
Total	= $tr - 1$	= $6 \times 6 - 1$	= 35

Note that in Latin Square, the number of treatments (t) equals the number of columns (c) equals the number of rows (r), only t will be used as divisor in the formula to find the Sums of Squares.

Sums of Squares

$$\text{C.F.} = \frac{(GT)^2}{t^2} = \frac{(551)^2}{6^2} = \boxed{8,433.3611}$$

Total Sum of Squares (ToSS). ToSS can be computed using the sequence of Treatment x Column, Treatment x Row, Column x Row, or Row x Column. In the example, Row x Column is used.

$$= \Sigma \Sigma (\text{RoCo})^2 - \text{C.F.} = (\text{Ro}_1\text{Co}_1)^2 + (\text{Ro}_1\text{Co}_2)^2 + \dots + (\text{Ro}_6\text{Co}_6)^2 - \text{C.F.}$$

$$= (17^2 + 15^2 + \dots + 17^2) - 8,433.3611 = 8,745.0000 - 8,433.3611 = \boxed{311.6389}$$

Treatment Sum of Squares (TrSS)

$$= \frac{\Sigma T^2}{t} - \text{C.F.} = \frac{T_1^2 + T_2^2 + \dots + T_6^2}{t} - \text{C.F.}$$

$$= \frac{105^2 + 94^2 + \dots + 86^2}{6} - 8,433.3611 = 8,535.8333 - 8,433.3611 = \boxed{102.4722}$$

Column Sum of Squares (CoSS)

$$= \frac{\Sigma \text{Co}^2}{t} - \text{C.F.} = \frac{\text{Co}_1^2 + \text{Co}_2^2 + \dots + \text{Co}_6^2}{t} - \text{C.F.}$$

$$= \frac{97^2 + 105^2 + \dots + 94^2}{6} - 8,433.3611 = 8,511.5000 - 8,433.3611 = \boxed{78.1389}$$

Row Sum of Squares (RSS)

$$= \frac{\Sigma \text{Ro}^2}{t} - \text{C.F.} = \frac{\text{Ro}_1^2 + \text{Ro}_2^2 + \dots + \text{Ro}_6^2}{t} - \text{C.F.}$$

$$= \frac{90^2 + 96^2 + \dots + 78^2}{6} - 8,433.3611 = 8,499.1667 - 8,433.3611 = \boxed{65.8056}$$

Error Sum of Squares (ESS)

The Error df for Latin Square, $(t - 1)(t - 2)$, when expanded is $t^2 - 3t + 2$. The term t^2 is the same as tr in CRD or RCBD. The term $3t$ refers to Squares of Treatments, Squares of Columns, and Squares of Rows. Therefore, the formula to compute Error SS for Latin Square is:

$$= \Sigma \Sigma (\text{TR})^2 - \frac{\Sigma T^2}{t} - \frac{\Sigma \text{Co}^2}{t} - \frac{\Sigma \text{Ro}^2}{t} + 2 \text{ C.F.}$$

Since all these values have been computed as shown above, the final values are:

$$= 8,745.0000 - 8,535.8333 - 8,511.5000 - 8,499.1667 + 2 \times 8,433.3611 = \boxed{65.2222}$$

Mean Squares

$$\text{Row Mean Square (RoMS)} = \frac{\text{RoSS}}{\text{Ro df}} = \frac{65.8056}{5} = \boxed{13.1611}$$

$$\text{Column Mean Square (CoMS)} = \frac{\text{CoSS}}{\text{Co df}} = \frac{78.1389}{5} = \boxed{15.6278}$$

$$\text{Treatment Mean Square (TrMS)} = \frac{\text{TrSS}}{\text{Tr df}} = \frac{102.4722}{5} = \boxed{20.4944}$$

$$\text{Error Mean Square (EMS)} = \frac{\text{ESS}}{\text{E df}} = \frac{65.2222}{20} = \boxed{3.2611}$$

F-computed

$$\text{Row F-computed (RoFc)} = \frac{\text{RoMS}}{\text{EMS}} = \frac{13.1611}{3.2611} = \boxed{4.04}$$

$$\text{Column F-computed (CoFc)} = \frac{\text{CoMS}}{\text{EMS}} = \frac{15.6278}{3.2611} = \boxed{4.79}$$

$$\text{Treatment F-computed (TrFc)} = \frac{\text{TrMS}}{\text{EMS}} = \frac{20.4944}{3.2611} = \boxed{6.28}$$

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Row	5	65.8056	13.1611	4.04 *	4.10	2.71
Column	5	78.1389	15.6278	4.79 **	4.10	2.71
Treatment	5	102.4722	20.4944	6.28 **	4.10	2.71
Error	20	65.2222	3.2611			
Total	35	311.6389				

$$\text{C.V.} = \frac{\sqrt{(\text{EMS})}}{\text{Mean}} \times 100 = \frac{\sqrt{3.2611}}{15.3} \times 100 = \boxed{11.8\%}$$

Comparing Latin Square with CRD and RCBD, a number of observations would be noted.

- Using the same set of data, the treatment means in CRD and RCBD are declared significantly different while in LS design, they are highly significant.
- The analysis also showed significant differences among columns and among rows.
- The addition of columns and rows instead of replication alone resulted in further reduction of the residual error mean square from 5.2411 in RCBD to 3.2611 in LS.
- There was a reduction in C.V. from 15.0% in RCBD to 11.8% in Latin Square.
- The Relative Efficiency of LS over CRD is computed using the formula:

$$\text{R.E.(CRD)} = \frac{E_r + E_c + r(t-1)E_e}{(t+1)E_e} = \frac{13.1611 + 15.6278 + (6-1) \times 5.2411}{(6+1) \times 5.2411} = \boxed{1.50}$$

- The relative efficiency of a LS design as compared to a RCB design can be computed in two ways—when rows are considered as blocks, and when columns are considered as blocks, of the RCB design.

$$\text{R.E.(RCB, row)} = \frac{E_r + (t-1)E_e}{t \times E_e} = \frac{13.1611 + (6-1) \times 5.2411}{6 \times 5.2411} = \boxed{1.25}$$

$$\text{R.E.(RCB, col)} = \frac{E_c + (t-1)E_e}{t \times E_e} = \frac{15.6278 + (6-1) \times 5.2411}{6 \times 5.2411} = \boxed{1.33}$$

- Analysis of Relative Efficiency of LS over CRD showed a 50% increase in experimental precision.
- Compared to RCBD, there was in 25% increase in precision when rows are considered as blocks; and 33% when columns are considered as block.
- Over-all, the use of rows and columns increased experimental precision.

Mean Comparison using LSD

$$s_d = \sqrt{\frac{2 \times 3.2611}{6}} = \boxed{1.0620}$$

$$LSD_{0.05} = s_d \times t_{0.05, 20 \text{ df}} = 1.0620 \times 2.086 = \boxed{2.2}$$

$$LSD_{0.01} = s_d \times t_{0.01, 20 \text{ df}} = 1.0620 \times 2.845 = \boxed{3.0}$$

Since ANOVA declared that the Treatment means are highly significantly different, either LSD 5% or LSD 1% may be used to compare Treatment means depending on what level of confidence the researcher is interested in. With 6 treatments, 5 valid pair comparisons can be made. Comparing the check (Treatment 6) with the other treatments.

Treatment	Mean	Difference
Treatment 1	17.5 **	3.3
Treatment 2	15.7 ns	1.5
Treatment 3	16.8 *	2.7
Treatment 4	15.2 ns	1.0
Treatment 5	12.3 ns	1.9
Treatment 6 (check)	14.2	

With 99% confidence level, LSD 1% is applied which showed that only Treatment 1 is significantly different from the check. If LSD 5% is used (95% confidence level), Treatments 1 and 3 are detected to be significantly different from the check.

Mean Comparison using DMRT

The same procedure in computation of DMRT as in CRD and RCBD is followed but the Rp values will change because of the change in Error df.

When completed, the final DMRT table will be:

Treatment 1	17.5 a
Treatment 3	16.8 ab
Treatment 2	15.7 abc
Treatment 4	15.2 abcd
Treatment 6 (chk)	14.3 cde
Treatment 5	12.3 e

For the interpretation of results, Treatments 1 & 3 are significantly different from Treatments 6 & 5 because they do not have a common letter. Treatments 2 & 4 are significantly different from Treatment 5. Treatments 6 & 5 are not significantly different.

Comparing the check with the rest of the treatments, It is significantly different from Treatments 1 & 3 but not from the other treatments. If the data were on yield, Treatments 1 & 3 are significantly better than the check.

PROBLEM DATA

Analysis of variance is applicable if certain assumptions are met. Some of the conditions are implied while others are specified. In field experiments, it is implied that all plots are grown successfully and all necessary data are taken and recorded properly. In addition, it is specified that the data satisfy all the mathematical assumptions underlying the analysis of variance.

The term *problem data* is used for any set of data that does not satisfy the implied or stated conditions for a valid analysis of variance. There are two types of problem data:

- Missing data
- Data that violate some assumptions of the analysis of variance.

Missing Data

A missing data occurs when a valid observation is not available for anyone of the experimental units. Occurrence of missing data results in two major difficulties: loss of information and non-applicability of the standard analysis of variance.

Common causes of missing data

1. Improper treatment. Improper treatment is declared when an experiment has one or more experimental plots that do not receive the intended treatment. Examples are non-application, application of an incorrect dose, and wrong timing of application. However, if the improper treatment occurs in all the replications, the researcher may consider retaining the treatment but the experimental objectives will have to be changed.
2. Destruction of experimental plants. Most field experiments aim to maintain a perfect stand but this is not always the case. Poor germination, physical damage during crop culture, and pest damage are common causes of destruction of experimental plants. If the percentage of destroyed plants in a plot is relatively small, proper thinning may correct the problem so there is no need to declare missing data. However, if the percentage of destroyed plants is very high and no valid observation can be made for that particular plot, missing data should be declared.

Destruction of plants in a plot directly caused by the treatment applied should not be declared missing data. Example is the application of different concentrations of an insecticide. If a high dosage has caused destruction of the plants, the corresponding plot data should be gathered and entered.

3. Loss of harvested samples. There are data that cannot be taken directly in the field and plant samples are transported from the field to the work area or laboratory. Examples are leaf area when using leaf area reader, 100-grain yield, or protein content. There are cases when samples are lost along the way or misplaced such that recording of the correct data is not possible. In this case missing data should be declared.
4. Illogical data. These type of data are usually recognized after the data have been recorded and transcribed. Data may be considered illogical if their values are too extreme to be considered within the logical range of the normal behavior of the experimental materials. Common errors in illogical data are misread observation, incorrect transcription, and improper application of the sampling technique or the measuring instrument. It should be emphasized that data a researcher suspects to be illogical should not be treated as missing data simply because they do not conform to the researcher's pre-conceived ideas or hypothesis. An observation considered to be illogical because it falls outside the researcher's expected range of values can be judged missing only if it can be shown to be caused by an error.

Note that missing data technique is not necessary for CRD experiments. In case one or several data are missing in a CRD experiment, just use the unequal number of replication technique in the analysis. Therefore, this technique is normally used for RCBD experiments.

Missing data technique for RCBD

$$X = \frac{rB_o + tT_o - G_o}{(r-1)(t-1)}$$

where:

X = estimate of the missing value

t = number of treatments

r = number of replications

B_o = total of observed values of the block containing the missing value

T_o = total of observed values of the treatment containing the missing value

G_o = initial Grand Total

The missing data is replaced by the computed value of X and the usual computational procedures for the analysis of variance are applied with some adjustments. Using the data in the following example with 4 replications and assuming the value for Treatment 6 Replication 4 is missing, the raw data table will be as follows:

TREATMENT	REP 1	REP 2	REP 3	REP 4	TOTAL
Treatment 1	17.2	19.7	17.4	17.5	71.8
Treatment 2	18.4	13.6	19.2	10.9	62.1
Treatment 3	17.8	21.8	18.4	13.8	71.8
Treatment 4	16.5	22.4	14.2	11.8	64.9
Treatment 5	15.4	12.1	12.4	13.5	53.4
Treatment 6	13.3	14.9	12.8		<u>41.0</u>
TOTAL	98.6	104.5	94.4	<u>67.5</u>	365.0

Computing for the missing value:

$$X = \frac{4(67.5) + 6(41.0) - 365.0}{(4-1)(6-1)} = 10.1$$

The value 10.1 is placed in the missing data and the totals recalculated. The new raw data table will be:

TREATMENT	REP 1	REP 2	REP 3	REP 4	TOTAL
Treatment 1	17.2	19.7	17.4	17.5	71.8
Treatment 2	18.4	13.6	19.2	10.9	62.1
Treatment 3	17.8	21.8	18.4	13.8	71.8
Treatment 4	16.5	22.4	14.2	11.8	64.9
Treatment 5	15.4	12.1	12.4	10.7	53.4
Treatment 6	13.3	14.9	12.8	<u>10.1</u>	<u>51.1</u>
TOTAL	98.6	104.5	94.4	<u>77.6</u>	<u>375.1</u>

Because one estimated value added in the data, correction factor for **bias B** needs to be computed based on the formula:

$$B = \frac{[Bo - (t - 1)X]^2}{t(t - 1)} = \frac{[67.5 - (6 - 1)10.1]^2}{6(6 - 1)} = 9.8231$$

This value will be subtracted from the Treatment and Total Sum of Squares. The error df and total df will be reduced by 1. By applying the normal analysis of variance for RCBD, the following results are obtained:

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	3	66.7213	22.2404	3.21 ns	5.56	3.34
Treatment	5	88.0939	17.6168	2.54 ns	4.59	2.96
Error	14	97.0712	6.9337			
TOTAL	22	251.8864				

$$C.V. = \frac{\sqrt{EMS}}{\text{Mean}} \times 100 = \frac{\sqrt{6.9337}}{15.63} \times 100 = 16.8\%$$

If there were no missing value and the data will be analyzed using normal method, the ANOVA will be:

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	3	66.7213	22.2404	3.44 *	5.42	3.29
Treatment	5	97.9171	19.5834	3.03 *	4.56	2.90
Error	15	97.0712	6.4714			
TOTAL	23	261.7096				

$$C.V. = \frac{\sqrt{EMS}}{\text{Mean}} \times 100 = \frac{\sqrt{6.4714}}{15.63} \times 100 = 16.3\%$$

Note that when the significance is very near the critical level at 5% probability, the use of missing value technique may result to non-significance of treatment mean differences. This is brought about by the reduction of Treatment Sum of Squares (due to subtraction of Bias B) and the reduction in the number of Error df (resulting to higher Error MS).

In case two or more values are missing, a decision has to be made whether the use of missing value technique is appropriate. As indicated in the previous example, a bias B is computed for every missing value which is the subtracted from the Treatment SS. This will cause the Treatment SS to be reduced drastically as more missing values are computed. At the same time, the Error df is reduced by the number of missing values resulting to higher Error MS. Since Treatment MS (reduced in value) is divided by the Error MS (increased in value), the resulting Treatment Fc will become smaller and smaller as the number of missing values increases.

There are other options that can be used instead of missing value technique in case two or more values are missing.

- If the missing values are found in the same block, it will be advisable to remove that block (provided there are more than two blocks in the experiment). NOTE: it is not allowed to

move one value from one block to another and then remove the block with the highest number of missing values.

- If the missing values are from the same treatment, that treatment can be excluded from the analysis (provided the removal will not affect the over-all objective of the experiment).

In the case of two missing values in RCBD experiment, the same procedure as in single missing value technique is used but with iteration or repetition of the procedure until the final acceptable answer is obtained. Consider a case with two missing values (T_1R_1 and T_6R_4) using the following table:

TREATMENT	REP 1	REP 2	REP 3	REP 4	TOTAL
Treatment 1		19.7	17.4	17.5	54.6
Treatment 2	18.4	13.6	19.2	10.9	62.1
Treatment 3	17.8	21.8	18.4	13.8	71.8
Treatment 4	16.5	22.4	14.2	11.8	64.9
Treatment 5	15.4	12.1	12.4	13.5	53.4
Treatment 6	13.3	14.9	12.8		41.0
TOTAL	81.4	104.5	94.4	67.5	347.8

Since it is not possible to remove any of the blocks or any of the treatments, the missing value technique will be used. The procedure requires computing for a missing value one at a time. First, assign a temporary number to one of the missing values. Although any value may be used, it is advisable to use a value as close as possible to the mean of that treatment. Very extreme value (very high or very low) may lead to longer iteration and computation before achieving the accurate missing value. In the example, a temporary value will be assigned to T_1R_1 . The average of all values in Treatment 1 is 18.2 and this value will be used. The first temporary table will be:

TREATMENT	REP 1	REP 2	REP 3	REP 4	TOTAL
Treatment 1	<u>18.2</u>	19.7	17.4	17.5	72.8
Treatment 2	18.4	13.6	19.2	10.9	62.1
Treatment 3	17.8	21.8	18.4	13.8	71.8
Treatment 4	16.5	22.4	14.2	11.8	64.9
Treatment 5	15.4	12.1	12.4	13.5	53.4
Treatment 6	13.3	14.9	12.8		41.0
TOTAL	99.6	104.5	94.4	67.5	366.0

Compute for the missing value for T_6R_4 :

$$X_1 = \frac{4(67.5) + 6(41.0) - 366.0}{(4-1)(6-1)} = \boxed{10.0}$$

Put the value 10.0 to T_6R_4 and remove the first temporary value of 18.2 from T_1R_1 .

TREATMENT	REP 1	REP 2	REP 3	REP 4	TOTAL
Treatment 1		19.7	17.4	17.5	54.6
Treatment 2	18.4	13.6	19.2	10.9	62.1
Treatment 3	17.8	21.8	18.4	13.8	71.8
Treatment 4	16.5	22.4	14.2	11.8	64.9
Treatment 5	15.4	12.1	12.4	13.5	53.4
Treatment 6	13.3	14.9	12.8	<u>10.0</u>	51.0
TOTAL	81.4	104.5	94.4	77.5	357.8

Consider 10.0 as the new temporary value, re-compute the totals and compute for the second missing value (T_1R_1) using the same procedure.

$$X_2 = \frac{4(81.4) + 6(54.6) - 357.8}{(4-1)(6-1)} = \boxed{19.7}$$

Put the value 19.7 in T_1R_1 and remove the value 10.0 from T_6R_4 , re-compute the totals and apply again the missing value formula:

TREATMENT	REP 1	REP 2	REP 3	REP 4	TOTAL
Treatment 1	<u>19.7</u>	19.7	17.4	17.5	74.3
Treatment 2	18.4	13.6	19.2	10.9	62.1
Treatment 3	17.8	21.8	18.4	13.8	71.8
Treatment 4	16.5	22.4	14.2	11.8	64.9
Treatment 5	15.4	12.1	12.4	13.5	53.4
Treatment 6	13.3	14.9	12.8		41.0
TOTAL	101.1	104.5	94.4	67.5	367.5

$$X_1 = \frac{4(67.5) + 6(41.0) - 367.5}{(4-1)(6-1)} = \boxed{9.9}$$

Continue replacing the other value, compute the totals and apply the missing value formula until both values are not changing anymore. Normally, after three iterations, the two missing values will settle and will not change. In the example, the final missing values are 19.7 for T_1R_1 and 9.9 for T_6R_4 . The final table will be:

TREATMENT	REP 1	REP 2	REP 3	REP 4	TOTAL
Treatment 1	<u>19.7</u>	19.7	17.4	17.5	74.3
Treatment 2	18.4	13.6	19.2	10.9	62.1
Treatment 3	17.8	21.8	18.4	13.8	71.8
Treatment 4	16.5	22.4	14.2	11.8	64.9
Treatment 5	15.4	12.1	12.4	13.5	53.4
Treatment 6	13.3	14.9	12.8	<u>9.9</u>	50.9
TOTAL	101.1	104.5	94.4	77.4	377.4

The data can now be computed using the usual analysis of variance. After computing the sums of squares, the incomplete ANOVA table will be:

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	3	72.6483				
Treatment	5	112.0150				
Error	13	93.1817				
TOTAL	21	277.8450				

Two correction bias B_1 & B_2 should be computed and subtracted from the Treatment SS and Total SS. Likewise, 2 df should be deducted from the Error df and Total df.

$$B_1 = \frac{[B_{10} - (t-1)X_1]^2}{t(t-1)} = \frac{[81.4 - (6-1)19.7]^2}{6(6-1)} = \boxed{9.7470}$$

$$B_2 = \frac{[B_{20} - (t-1)X_2]^2}{t(t-1)} = \frac{[67.5 - (6-1)9.9]^2}{6(6-1)} = \boxed{10.8000}$$

After deducting these two B values from the Treatment and Total SS and 2 df from Error and Total df, the ANOVA table will be:

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	3	72.6483	24.2161	3.38 ns	5.74	3.41
Treatment	5	91.4680	18.2936	2.55 ns	4.86	3.02
Error	13	93.1817	7.1678			
TOTAL	21	257.2980				

$$C.V. = \frac{\sqrt{EMS}}{\text{Mean}} \times 100 = \frac{\sqrt{7.1678}}{15.7} \times 100 = 17.0\%$$

If there is a need to use the missing value technique in an experiment with more than two (2) missing values, the procedure is the same as in two missing values with 2 or more temporary values being assigned and one missing value being calculated at a time. The computed missing value is placed in the corresponding missing value and the next missing value is computed, and so on.

Note that the number of missing values will be equal to:

- the number of correction factors for bias B to be computed and deducted from the treatment SS and Total SS.
- the number of degrees of freedom to be deducted from Error and Total df.

Data that violate some assumptions of the analysis of variance

Interpretation of the analysis of variance is valid only when certain mathematical assumptions concerning the data are met. When these assumptions are not met, then the analysis should not be considered valid. These assumptions are:

1. Non-additive effects. The effects of two factors such as treatment and replication are considered additive if the effect of one factor remains constant over all levels of the other factor, that is, if the treatment effect remains constant for all replications and replication effect remains constant for all treatments, the effects of treatment and replication are additive.
2. Non-independence of errors. The assumption requires that the error (deviation) of an observation is not related to, or dependent upon, that of another.
3. Heterogeneity (or non-uniformity) and non-normality of variance. There are two types:
 - Where the variance is functionally related to the mean. This is normally associated with data whose distribution does not follow the normal curve. Examples are count data such as number of infested plants per plot, or number of lesions per leaf.
 - Where there is no functional relationship between the variance and the mean. This occurs in an experiment where, due to the nature of the treatments tested, some treatments have errors that are substantially higher (or lower) than others.

One common method used to correct data that violate the assumptions for a valid analysis of variance is data transformation. It is the most appropriate remedial measure for variance heterogeneity where the variance and the mean are functionally related. With the technique, the original data are converted into a new scale resulting in a new data set that is expected to satisfy the condition of homogeneity (or uniformity) of variance. Because a common transformation scale is applied to all observations, the comparative values between treatments are not altered and comparisons between them remain valid.

The appropriate data transformation to be used depends on the specific type of relationship between the variance and the mean. Three most commonly used transformation for data in agricultural research are:

- Logarithmic transformation
- Square root transformation
- Arcsine transformation

Logarithmic transformation

It is most appropriate for data where the standard deviation is proportional to the mean or where the effects are multiplicative. These conditions are usually found in data that are whole numbers and cover a wide range of values. Examples are number of insects egg masses per plant, or number of insects per plant.

If the data set involves a wide range of values, take the logarithm of each and every component of the data set. If the data involve small values (less than 10), $\log(Y + 1)$ should be used instead of $\log Y$ where Y is the original value.

A hypothetical example involving insect count on 9 treatments is shown below.

Treatment	REP 1	REP 2	REP 3	TOTAL	Trt Mean
T1	10	17	8	35	12
T2	22	45	50	117	39
T3	15	280	12	307	102
T4	185	98	250	533	178
T5 (control)	755	102	512	1369	456
T6	400	750	231	1381	460
T7	450	655	845	1950	650
T8	1214	758	550	2522	841
T9	3547	2370	1578	7495	2498
TOTAL	6598	5075	4036	15709	582

When the original set of data is analyzed, the following results are obtained:

SV	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	2	368,996.0741	184,498.0370	1.27	6.23	3.63
Treatment	8	14,364,465.4074	1,795,558.1759	12.38	3.89	2.59
Error	16	2,320,406.5926	145,025.4120			
TOTAL	26	17,053,868.0741				

C.V. = 65.5 %

It will be noted that although the Treatment Means are declared highly significantly different as shown by higher Fc compared to Ft 1% ($12.38 > 3.89$), the C.V. is relatively high which makes the results questionable or not reliable. This may happen because data on count, especially those involving very low and very high do not follow the normal distribution curve. To correct or minimize the problem, logarithmic transformation is done:

Treatment	REP 1	REP 2	REP 3	TOTAL	Trt Mean
T1	1.00	1.23	0.90	3.13	1.04
T2	1.34	1.65	1.70	4.69	1.56
T3	1.18	2.45	1.08	4.70	1.57
T4	2.27	1.99	2.40	6.66	2.22
T5 (control)	2.88	2.01	2.71	7.60	2.53
T6	2.60	2.88	2.36	7.84	2.61
T7	2.65	2.82	2.93	8.40	2.80
T8	3.08	2.88	2.74	8.70	2.90
T9	3.55	3.37	3.20	10.12	3.37
TOTAL	20.55	21.28	20.02	61.85	2.29

When analysis of variance is done on the transformed data, the following results are obtained:

SV	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	2	0.0887	0.0444	0.35	6.23	3.63
Treatment	8	13.7270	1.7159	13.66	3.89	2.59
Error	16	2.0104	0.1257			
TOTAL	26	15.8261				

C.V. = 15.5%

Proper transformation of data resulted to the reduction of coefficient of variation from 65.5% to 15.5% which is within the acceptable limit.

Since the treatment means are highly significantly different, the means may be compared using either LSD or DMRT. Since there is a control, the 8 treatment means can be compared with the control using LSD. In this case the LSD should be done on the transformed values and then the corresponding significance transferred to the original Treatment Means. To illustrate, the transformed Treatment Means are:

Treatment	Mean
T ₁	1.04
T ₂	1.56
T ₃	1.57
T ₄	2.22
T ₅ (control)	2.53
T ₆	2.61
T ₇	2.80
T ₈	2.90
T ₉	3.37

$$\text{Compute for LSD 5\%} = \sqrt{\frac{2 \times 0.1257}{3}} \times 2.12 = \boxed{0.61}$$

The LSD value should be compared to the absolute difference between any transformed treatment mean and the control mean. If the difference between a Treatment Mean and the Control Mean is equal to or greater than 0.61, that Treatment Mean is declared significantly different from the Control Mean so a star (*) is placed after the mean value. Any transformed Treatment Mean whose difference from the Control Mean is lower than 0.61 will be declared not significantly different and should be followed by "ns". After comparing all the transformed Treatment Means, T₁, T₂, and T₃ were found to be significantly lower than the Control Mean in terms of number of insects per plot. Likewise, T₉ was significantly higher than the Control. All other transformed Treatment Means are declared not significantly different from the transformed Control Mean. The transformed means will have the following stars:

Treatment	Mean
T ₁	1.04 *
T ₂	1.56 *
T ₃	1.57 *
T ₄	2.22 ns
T ₅ (control)	2.53
T ₆	2.61 ns
T ₇	2.80 ns
T ₈	2.90 ns
T ₉	3.37 *

The "*" and "ns" notations should now be transferred to the original Treatment Means. The final Treatment Means and LSD notations will be:

Treatment	Insect Count Mean
T ₁	12 *
T ₂	39 *
T ₃	102 *
T ₄	178 ns
T ₅ (control)	456
T ₆	460 ns
T ₇	650 ns
T ₈	841 ns
T ₉	2498 *

Square root transformation

It is appropriate for data consisting of small whole numbers, e.g. number of infested plants per plot, number of weeds per plot, number of insects caught in a trap. These data tend to be proportional to the mean. Square root transformation is also appropriate for percentage data whose range is between 0 and 30% or between 70 and 100%.

If most of the values are small (less than 10), especially where zeroes are present, use the formula

$$Y' = \sqrt{Y + 0.5}$$

where Y' is the transformed value, and

Y is the original value

A hypothetical example of 9 treatments with percent values ranging from 6.9% to 27.6% will be used.

Treatment	REP 1	REP 2	REP 3	TOTAL	MEAN
1	27.4	25.6	17.3	70.3	23.4
2	22.1	18.4	25.6	66.1	22.0
3	6.6	10.4	6.2	23.2	7.7
4	25.4	22.1	14.3	61.8	20.6
5	7.4	6.9	7.2	21.5	7.2
6	19.6	25.2	5.3	50.1	16.7
7	20.2	15.6	11.7	47.5	15.8
8	27.6	20.4	15.8	63.8	21.3
9	10.2	17.4	25.4	53.0	17.7
TOTAL	166.5	162	128.8	457.3	16.9

Since the range of values falls to the category 0-30%, square root transformation will be used. There is no zero value so there is no need to add 0.5 to all the values. The square root transformation values are shown below.

Treatment	REP 1	REP 2	REP 3	TOTAL	MEAN
1	5.23	5.06	4.16	14.45	4.82
2	4.70	4.29	5.06	14.05	4.68
3	2.57	3.22	2.49	8.28	2.76
4	5.04	4.70	3.78	13.52	4.51
5	2.72	2.63	2.68	8.03	2.68
6	4.43	5.02	2.30	11.75	3.92
7	4.49	3.95	3.42	11.86	3.95
8	5.25	4.52	3.97	13.75	4.58
9	3.19	4.17	5.04	12.40	4.13
TOTAL	37.63	37.56	32.91	108.10	4.00

If the original data were analyzed, the results will be:

SV	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	2	94.2141	47.1070	1.51	6.23	3.63
Treatment	8	846.9363	105.8670	3.40	3.89	2.59
Error	16	498.8126	31.1758			
TOTAL	26	1,439.9630				

C.V. = 33.0%

If the data were transformed using square root, the ANOVA will be:

SV	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	2	1.6264	0.8132	1.69	6.23	3.63
Treatment	8	15.1325	1.8916	3.93	3.89	2.59
Error	16	7.7051	0.4816			
TOTAL	26	24.4640				

C.V. = 17.3%

Similar to the results in logarithmic transformation, square root transformation reduced C.V. from 33.0% to 17.3% and the resulting Fc was detected to be highly significant.

Arcsine transformation

An arcsine or angular transformation is appropriate for data on proportion, data obtained from a count and data expressed as decimal fractions or percentages. Note that percentages based on direct measurement, e.g. % protein content reading from a machine, should not be transformed using arcsine.

Not all percentage data need to be transformed and, even if they do, not all of them require arcsine transformation. The following rules must be used:

- For percentage data lying between 30 & 70%, there is no need to do any transformation. The data can be analyzed directly.
- For percentage data lying between 0 and 30%, or between 70 and 100%, but not both, the square root transformation should be used.
- For percentage data lying between 0 and 100%, arcsine transformation should be used.

When there is a zero value and the total number of observations is less than 50, the 0 value should be replaced using the formula:

$$Y = \frac{1}{4n}$$

where n is the number of units upon which the percentage data was based.

For example, in an experiment on insect damage with 20 plants per plot, 0% damage should be replaced by:

$$Y = \frac{1}{4 \times 20} = 0.0125 \text{ or } \boxed{1.25\%}$$

The arcsine transformation of this value (= 6.42) should be used instead of the arcsine transformation of 0 (= 0.00).

Likewise, if there is a 100% value and the number of units is less than 50, it should be replaced by the formula:

$$Y = 1 - \frac{1}{4n}$$

In the same example, a 100% value should be replaced by:

$$Y = 1 - \frac{1}{4 \times 20} = 0.9875 \text{ or } \boxed{98.75\%}$$

The arcsine transformation of this value (= 83.58) should be used instead of the arcsine transformation of 100% (= 90.00).

Arcsine transformation values are available in most statistical books. Transformation can also be computed using normal scientific calculator (with arcsine function), spreadsheet software (EXCEL, LOTUS) and statistical package in computer (SAS, SPSS, MSTAT, IRRISTAT). The formula to be used is:

$$Y' = \arcsin \sqrt{Y}$$

where Y' = transformed value

Y = original value in decimal (not percent) format.

If the values are expressed in percentage (75%, 36%, etc.), they should be converted first to decimals (0.75, 0.36, etc.) before extracting the square root and then the corresponding arcsine value is computed.

The arcsine value is expressed in two formats:

- Radians = maximum value (for 100%) is 1.5707963267949 (half the value of π [π])
- Degrees = maximum value (for 100%) is 90.00

Arcsine transformation tables found in statistical books are normally expressed in degrees while values in statistical softwares are expressed in radians. To convert radians to degrees, use the formula:

$$Y' = Y \times (180 / \pi)$$

where

Y' = value in degrees

Y = value in radians

π = value of pi or 3.14159265358979

Note that arcsine transformed data, whether in radians or degree format, will result to basically the same Fc value although the sum of squares and mean squares values will differ.

If possible, it is advisable to directly compute arcsine transformation value rather than getting from table because computation will result to the actual or nearest arcsine value. Values from the table are approximations and have been rounded-off to the two nearest decimals.

Treatment	REP 1	REP 2	REP 3	TOTAL	MEAN
1	93.4	50.6	86.2	230.2	76.7
2	78.6	58.4	84.5	221.5	73.8
3	66.6	85.4	76.2	228.2	76.1
4	65.3	38.5	44.3	148.1	49.4
5	22.4	65.3	37.4	125.1	41.7
6	48.6	56.7	74.3	179.6	59.9
7	88.6	55.2	67.3	211.1	70.4
8	37.6	60.7	55.8	154.1	51.4
9	35.2	25.6	75.4	136.2	45.4
TOTAL	536.3	496.4	601.4	1634.1	60.5

When analyzed directly without transformation, the following results will be obtained:

SV	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	2	624.2600	312.1300	1.01	6.23	3.63
Treatment	8	4,710.5600	588.8200	1.91	3.89	2.59
Error	16	4,932.6267	308.2892			
TOTAL	26	10,267.4467				

C.V. = 29.0%

Using arcsine transformation (expressed in degrees), the following values will be obtained:

Treatment	REP 1	REP 2	REP 3	TOTAL	MEAN
1	75.11	45.34	68.19	188.65	62.9
2	62.44	49.84	66.82	179.10	59.7
3	54.70	67.54	60.80	183.03	61.0
4	53.91	38.35	41.73	133.99	44.7
5	28.25	53.91	37.70	119.86	40.0
6	44.20	48.85	59.54	152.59	50.9
7	70.27	47.98	55.12	173.37	57.8
8	37.82	51.18	48.33	137.33	45.8
9	36.39	30.40	60.27	127.05	42.4
TOTAL	463.09	433.39	498.49	1394.97	51.7

And the final ANOVA will be:

SV	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	2	236.1093	118.0547	0.97	6.23	3.63
Treatment	8	1870.6185	233.8273	1.93	3.89	2.59
Error	16	1939.1386	121.1962			
TOTAL	26	4045.8664				

C.V. = **21.3%**

With proper transformation, the C.V. has been reduced from 29.0% to 21.3% although both results did not show significant differences among the Treatment Means.

Using arcsine transformation (expressed in radians), the following values will be obtained:

Treatment	REP 1	REP 2	REP 3	TOTAL	MEAN
1	1.31	0.79	1.19	3.29	1.10
2	1.09	0.87	1.17	3.13	1.04
3	0.95	1.18	1.06	3.19	1.06
4	0.94	0.67	0.73	2.34	0.78
5	0.49	0.94	0.66	2.09	0.70
6	0.77	0.85	1.04	2.66	0.89
7	1.23	0.84	0.96	3.03	1.01
8	0.66	0.89	0.84	2.40	0.80
9	0.64	0.53	1.05	2.22	0.74
TOTAL	8.08	7.56	8.70	24.35	0.90

When analyzed, the following results will be obtained:

SV	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	2	0.0719	0.0360	0.97	6.23	3.63
Treatment	8	0.5698	0.0712	1.93	3.89	2.59
Error	16	0.5907	0.0369			
TOTAL	26	1.2324				

C.V. = **21.3%**

Note that when using radians as units, the SS and MS values will be reduced drastically compared to when using degrees. However, the Fc and C.V. will be exactly the same.

MULTIPLE FACTOR EXPERIMENTS

The basic designs can be used when dealing with a single factor or treatment. However, some experiments require two or more factors or treatments to be analyzed together. As a general rule, it is always advisable to deal with simple experiments rather than complex ones, however, experiments involving two or more factors are conducted when the response to the factor of interest is expected to differ under different levels of the other factor, in short, an interaction occurs between the two factors. In dealing with interaction effects, the following points should be noted:

- In interaction, the effect between two factors can be measured only if the two factors are tested together in the same experiment.
- When interaction is absent, the simple effect of a factor is the same for all levels of the other factor(s) and equals the main effect.
- When interaction is present, the simple effect of a factor changes as the level of the other factor changes. Consequently, the main effect is different from the simple effects.

Based on the above points when dealing with interaction effects, the following points must be observed when interaction effect between two factors is present:

- The simple effects and not the main effects should be examined.
- The result from a single-factor experiment is applicable only to the particular level in which the other factors were maintained in the experiment and there can be no generalization of the result to cover any other levels.

FACTORIAL

Factorial, whether in CRD or RCBD, is applicable when an experiment is dealing with two factors to be evaluated at the same time. The other assumption is that both factors (A and B) and the interaction between the two (A x B) are of equal importance. Factorial can also be considered as simple CRD or RCBD if the factor combination will be considered as single treatment. After analysis, the treatment effect is partitioned into three meaningful components: due to A, due to B and due to AxB.

FACTORIAL IN COMPLETELY RANDOMIZED DESIGN

Randomization

The randomization procedure in factorial in CRD is very similar to CRD if the factor combination will be considered as single treatment. The steps in randomization are as follows:

1. Determine the number of A factor x B factor x Replication combinations in the whole experiment. In the example, there are 48 combinations (3 levels of A, 4 levels of B, and 4 replications).
2. Assign a plot number to each combination consecutively.
3. Assign the factor combinations to the experimental plots using a randomization scheme.
 - A. Using the table of random numbers
 1. With eyes closed, point the finger to the table of random numbers.
 2. Copy the middle 3 digits of all the consecutive digit numbers corresponding to the number of factor combinations.
 3. Rank the 3-digit numbers from 1 to the highest number of combinations.
 4. Arrange the 3-digit numbers consecutively from lowest to highest but be sure to carry the assigned rank. The sequence of the ranks will now serve as the randomized numbers.
 5. Assign the ranks corresponding to the number of combinations of the whole experiment.

B. Using draw lots

1. Prepare pieces of papers corresponding to the number of A factor x B factor x replication combinations of the whole experiment.
2. Write the combinations in each of the paper.
3. Mix the papers thoroughly in a container.
4. Without looking in the box, draw a paper and assign the combination on the first experimental unit of the whole experiment.
5. Without returning the paper previously drawn, continue drawing individual papers and assign the corresponding factor combination name until all the factor combinations have been assigned to the whole experiment.

Using the randomization discussed above, a possible lay-out of a 3 x 4 factorial with 4 replications is as follows:

$A_1B_1R_1$	$A_1B_3R_1$	$A_1B_2R_2$	$A_2B_3R_4$
$A_2B_3R_3$	$A_2B_3R_1$	$A_2B_3R_2$	$A_1B_3R_2$
$A_1B_4R_1$	$A_1B_2R_1$	$A_2B_1R_4$	$A_2B_2R_2$
$A_3B_3R_1$	$A_3B_2R_2$	$A_1B_2R_3$	$A_3B_2R_3$
$A_1B_2R_4$	$A_1B_4R_2$	$A_1B_1R_4$	$A_2B_1R_2$
$A_3B_4R_2$	$A_3B_1R_1$	$A_1B_1R_3$	$A_1B_3R_3$
$A_2B_1R_3$	$A_2B_2R_1$	$A_3B_4R_3$	$A_3B_1R_3$
$A_2B_2R_4$	$A_1B_3R_4$	$A_2B_4R_3$	$A_2B_1R_1$
$A_3B_2R_4$	$A_3B_4R_1$	$A_1B_4R_4$	$A_3B_3R_2$
$A_3B_1R_4$	$A_1B_4R_3$	$A_2B_2R_3$	$A_3B_1R_2$
$A_2B_4R_1$	$A_1B_1R_2$	$A_3B_2R_1$	$A_2B_4R_2$
$A_3B_3R_4$	$A_2B_4R_4$	$A_3B_3R_3$	$A_3B_4R_4$

Note that in CRD, the lay-out could be anything provided the conditions are similar in the whole area. However, it is recommended that the factor combinations are placed side by side and not too far from one another to be sure that the conditions are similar.

Equal number of replications

Consider an experiment based on the lay-out above involving two factors with four replications. Factor A has three levels while Factor B has four levels. The hypothetical data are as follows

FACTOR A	FACTOR B	REP I	REP II	REP III	REP IV	AxB TOTAL	AxB MEAN
A ₁	B ₁	28	29	23	22	102	25.5
A ₁	B ₂	25	24	27	24	100	25.0
A ₁	B ₃	26	28	31	27	112	28.0
A ₁	B ₄	23	24	17	23	87	21.8
A ₂	B ₁	27	31	20	27	105	26.2
A ₂	B ₂	16	20	22	20	78	19.5
A ₂	B ₃	17	23	16	25	81	20.2
A ₂	B ₄	19	18	17	18	72	18.0
A ₃	B ₁	30	29	23	33	115	28.8
A ₃	B ₂	16	14	16	23	69	17.2
A ₃	B ₃	18	16	20	25	79	19.9
A ₃	B ₄	17	20	10	29	76	19.0
TOTAL						1,076	22.4

To facilitate computation of Sums of Squares, prepare a two-way table for Factors A and B as shown below. Copy the AxB Totals from the upper table in the corresponding boxes in the lower table. A Totals and B Totals can be easily computed. The sums of A Totals and of B Totals should be equal to the Grand Total as a double check that the addition is correct.

A x B SUMMARY TABLE

	B ₁	B ₂	B ₃	B ₄	A TOTALS	A MEANS
A ₁	102	100	112	87	401	25.1
A ₂	105	78	81	72	336	21.0
A ₃	115	69	79	76	339	21.2
B TOTALS	322	247	272	235	1,076	
B MEANS	26.8	20.6	22.7	19.6		

Basic values

$$\begin{aligned}
 a &= \text{number or levels of Factor A} &= 3 \\
 b &= \text{number or levels of Factor B} &= 4 \\
 r &= \text{number of replications} &= 4 \\
 GT &= (A_1B_1R_1) + (A_1B_1R_2) + \dots + (A_3B_4R_4) = 26 + 29 + \dots + 29 &= 1076 \\
 \bar{X} &= \frac{(GT)}{abr} = \frac{1,076}{(3 \times 4 \times 4)} &= 22.4
 \end{aligned}$$

Degrees of Freedom

$$\begin{aligned}
 \text{Factor A} &= (a - 1) = (3 - 1) = 2 \\
 \text{Factor B} &= (b - 1) = (4 - 1) = 3 \\
 \text{A x B} &= (a - 1)(b - 1) = (3 - 1)(4 - 1) = 6 \\
 \text{Error} &= ab(r - 1) = 3 \times 4 \times (4 - 1) = 36 \\
 \text{Total} &= abr - 1 = 3 \times 4 \times 4 - 1 = 47
 \end{aligned}$$

Squares

$$\frac{\Sigma A^2}{br} = \frac{A_1^2 + A_2^2 + A_3^2}{br} = \frac{(401)^2 + (336)^2 + (330)^2}{4 \times 4} = \boxed{24,266.6250}$$

$$\frac{\Sigma B^2}{ar} = \frac{B_1^2 + B_2^2 + B_3^2 + B_4^2}{ar} = \frac{(322)^2 + (247)^2 + (272)^2 + (235)^2}{3 \times 4} = \boxed{24,491.8333}$$

$$\frac{\Sigma \Sigma (AB)^2}{r} = \frac{(A_1 B_1)^2 + (A_1 B_2)^2 + \dots + (A_3 B_4)^2}{r} = \frac{(102)^2 + (100)^2 + \dots + (76)^2}{3} = \boxed{24,843.5000}$$

$$\Sigma \Sigma \Sigma (ABR)^2 = (A_1 B_1 R_1)^2 + (A_1 B_1 R_2)^2 + \dots + (A_3 B_4 R_4)^2 = (26^2 + 29^2 + \dots + 29^2) = \boxed{25,404.0000}$$

$$C.F. = \frac{GT^2}{abr} = \frac{(1076)^2}{(3 \times 4 \times 4)} = \boxed{24,120.3333}$$

Sum of Squares

$$ToSS = \Sigma \Sigma \Sigma (ABR)^2 - C.F. = 25,404.0000 - 24,120.3333 = \boxed{1,283.6667}$$

$$ASS = \frac{\Sigma A^2}{br} - C.F. = 24,266.6250 - 24,120.3333 = \boxed{168.2917}$$

$$BSS = \frac{\Sigma B^2}{ar} - C.F. = 24,491.8333 - 24,120.3333 = \boxed{371.5000}$$

$$ABSS = \frac{\Sigma \Sigma (AB)^2}{r} - \frac{\Sigma A^2}{br} - \frac{\Sigma B^2}{ar} + C.F. = 24,843.5000 - 24,288.6250 - 24,491.98333 + 24,120.3333 = \boxed{183.3750}$$

$$ESS = \Sigma \Sigma \Sigma (ABR)^2 - \frac{\Sigma \Sigma (AB)^2}{r} = 25,404.0000 - 24,843.5000 = \boxed{560.5000}$$

Mean Squares

$$\text{Factor A Mean Square (AMS)} = \frac{ASS}{A \text{ df}} = \frac{168.2917}{2} = \boxed{84.1458}$$

$$\text{Factor B Mean Square (BMS)} = \frac{BSS}{B \text{ df}} = \frac{371.5000}{3} = \boxed{123.8333}$$

$$A \times B \text{ Mean Square (ABMS)} = \frac{ABSS}{AB \text{ df}} = \frac{183.3750}{6} = \boxed{30.5625}$$

$$\text{Error Mean Square (EMS)} = \frac{\text{ESS}}{\text{E df}} = \frac{560.5000}{36} = \boxed{15.5694}$$

F-computed

$$\text{Factor A F-computed (A Fc)} = \frac{\text{AMS}}{\text{EMS}} = \frac{84.1458}{15.5694} = \boxed{5.40}$$

$$\text{Factor B F-computed (B Fc)} = \frac{\text{BMS}}{\text{EMS}} = \frac{123.8333}{15.5694} = \boxed{7.95}$$

$$\text{A x B F-computed (A x B Fc)} = \frac{\text{ABMS}}{\text{EMS}} = \frac{30.5625}{15.5694} = \boxed{1.96}$$

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Factor A	2	166.2917	84.1458	5.40 **	5.25	3.26
Factor B	3	371.5000	123.8333	7.95 **	4.38	2.86
A x B	6	183.3750	30.5625	1.96 ns	3.35	2.36
Error	36	560.5000	15.5694			
TOTAL	47	1,283.6667				

$$\text{C.V.} = \frac{\sqrt{\text{EMS}}}{\text{Mean}} \times 100 = \frac{\sqrt{15.5694}}{22.4} \times 100 = \boxed{17.6\%}$$

In evaluating at the results of a multiple factor experiment, look first at the F-test of interaction. If A x B is significant together with the two main factors (A & B), the significance in the main factors becomes meaningless. This is because the effect of the different levels of one factor is affected by the different levels of the other factor (hence the word interaction). If the F-test of interaction is not significant while that of the main factor is significant, the discussion should center on the main factor.

The results showed highly significant differences among the different levels of A and different levels of B but no significant differences among the interaction. This indicates that two separate CRD experiment may have been conducted and still the general results will be the same. A careful analysis of the treatment (factor) and interaction means will show the following:

	A ₁	A ₂	A ₃	Factor B Mean
B ₁	25.5	26.2	28.8	26.8
B ₂	25.0	19.5	17.2	20.6
B ₃	28.0	20.2	19.8	22.7
B ₄	21.8	18.0	19.0	19.6
Factor A Mean	25.1	21.0	21.2	22.4

Evaluating Factor A means showed that A₁ gave the highest value. Among the different levels of Factor B, A₁ had the highest value in 3 B levels (B₂, B₃ & B₄). Looking at the A levels, B₁ was the highest in A₂ & A₃. That is an indication of no significant differences among interaction means: the trend of values of different A levels will more or less be the same in the different B levels, and

the trend of values of B levels will also be more or less the same in the different A levels. In conclusion, it is safe to say that among Factor A levels, A₁ was the highest regardless of B level, and vice-versa, among Factor B levels, B₁ was the highest regardless of A level. This again indicates that any level of B can be used to test the different levels of A and the trend will be the same; or any level of A can be used to test the different levels of B and the trend will be the same. If interaction was significant or highly significant, it is entirely different. This will be discussed later.

Since the levels of Factor A and levels of Factor B are both highly significantly different, LSD or DMRT should be done to determine which means in each Factor are different from each other. Since both differences are highly significant, the LSD or DMRT confidence level may be 5% or 1% depending on the degree of confidence the researcher wishes to use.

For multiple factor designs, the standard error of mean difference (s_d) will vary depending on the number of kinds or levels of the factors involved.

Mean Comparison using LSD

Factor A levels

$$s_d = \sqrt{\frac{(2MSE)}{b \times r}} = \sqrt{\frac{2 \times 15.5694}{4 \times 4}} = \boxed{1.395}$$

$$LSD_{0.05, 36 \text{ df}} = 2.031 \times 1.395 = \boxed{2.83}$$

$$LSD_{0.01, 36 \text{ df}} = 2.727 \times 1.395 = \boxed{3.80}$$

Any of the two LSD values may be used to compare the Treatment Means of the 3 Factor A levels. Since there are 3 treatments, only 2 valid pair comparisons may be made. Comparing A₁ with A₂, and A₁ with A₃ shows that A₁ is significantly different from A₂ and also from A₃ at both 1% and 5% probability level.

Factor B levels

$$s_d = \sqrt{\frac{(2MSE)}{a \times r}} = \sqrt{\frac{2 \times 15.5694}{3 \times 4}} = \boxed{1.611}$$

$$LSD_{0.05, 36 \text{ df}} = 2.031 \times 1.611 = \boxed{3.27}$$

$$LSD_{0.01, 36 \text{ df}} = 2.727 \times 1.611 = \boxed{4.39}$$

Any of the two LSD values will be used to compare the Treatment Means of the 3 Factor B levels. Since there are 4 treatments, only 3 valid pair comparisons may be made. Comparing B₁ with B₂, B₁ with B₃, and B₁ with B₄ shows that at 1% probability level, B₁ is highly significantly different from B₂, B₃ and B₄.

Mean Comparison using DMRT

If DMRT is desired, the same procedure as in previous example in RCBD may be used using the corresponding s_d values computed above.

To compare the 3 A treatment means:

$$s_d = 1.395$$

Error df = 36

P	r _p	R _p
2	2.87	2.83
3	3.02	2.98

Comparing the 3 means using the appropriate R_p value:

$A_1 = 25.1$ a
 $A_3 = 21.2$ b
 $A_2 = 21.0$ b

To compare the 4 $\underline{B}$ treatment means:

$s_d = 1.611$

Error df = 36

P	r_p	R_p
2	2.87	3.27
3	3.02	3.44
4	3.11	3.54

Comparing the 4 B means using the appropriate R_p value:

$B_1 = 26.8$ a
 $B_4 = 22.4$ b
 $B_2 = 20.6$ b
 $B_3 = 19.6$ b

Unequal number of replications

Consider an experiment based on the lay-out above involving two factors with four replications (4). Factor A has three levels while Factor B has four levels. However, there are three (3) factor combinations without values ($A_1B_1R_2$, $A_2B_4R_3$ & $A_3B_1R_1$). As in simple CRD, there is no need to compute for missing values. The hypothetical data are shown in the next page.

FACTOR A	FACTOR B	REP I	REP II	REP III	REP IV	TOTAL	MEAN
A_1	B_1	28		23	22	73	24.3
A_1	B_2	25	24	27	24	100	25.0
A_1	B_3	26	28	31	27	112	28.0
A_1	B_4	23	24	17	23	87	21.8
A_2	B_1	27	31	20	27	105	26.2
A_2	B_2	16	20	22	20	78	19.5
A_2	B_3	17	23	16	25	81	20.2
A_2	B_4	19	18		18	55	18.3
A_3	B_1		29	23	33	85	28.3
A_3	B_2	16	14	16	23	69	17.2
A_3	B_3	18	16	20	25	79	19.8
A_3	B_4	17	20	10	29	76	19.0
TOTAL						1,000	22.22

Again, a summary table for A and B is needed but note that some of the values are missing so in the computation of the means, the number of observations should be noted. In the above hypothetical data, all Factor A means are obtained by dividing each Factor A Total by 13 instead of 16 since there is no observation for $A_1B_1R_2$, $A_2B_4R_3$ and $A_3B_1R_1$. Likewise, Factor B₁ mean is computed by dividing Factor B Total by 14 instead of 16 since there are no values in $A_1B_1R_2$ and $A_3B_1R_1$.

A x B SUMMARY TABLE

	B ₁	B ₂	B ₃	B ₄	A TOTALS	A MEANS
A ₁	73	100	112	87	372	24.8
A ₂	105	78	81	55	319	21.3
A ₃	85	69	79	76	309	20.6
B TOTALS	263	247	272	218	1000	
B MEANS	26.3	20.6	22.7	19.8		

Basic values

$$\begin{aligned}
 a &= \text{number or levels of Factor A} &= 3 \\
 b &= \text{number or levels of Factor B} &= 4 \\
 r &= \text{number of replications} &= 4 \\
 GT &= (A_1B_1R_1) + (A_1B_1R_2) + \dots + (A_3B_4R_4) = 26 + 29 + \dots + 29 &= 1076 \\
 \bar{X} &= \frac{(GT)}{\Sigma ab_r} = \frac{1,000}{(3 \cdot 4 + \dots + 4)} &= 22.4
 \end{aligned}$$

Degrees of Freedom

$$\begin{aligned}
 \text{Factor A} &= (a - 1) = (3 - 1) = 2 \\
 \text{Factor B} &= (b - 1) = (4 - 1) = 3 \\
 \text{A x B} &= (a - 1)(b - 1) = (3 - 1)(4 - 1) = 6 \\
 \text{Error} &= \Sigma(ab - 1) = (3 - 1) + (4 - 1) + \dots + (4 - 1) = 33 \\
 \text{Total} &= (abr - 1) = 3 \times 4 \times 4 - 1 = 44
 \end{aligned}$$

Squares

$$\begin{aligned}
 \frac{\Sigma A^2}{\Sigma br_a} &= \frac{A_1^2}{br_{a1}} + \frac{A_2^2}{br_{a2}} + \frac{A_3^2}{br_{a3}} = \frac{(372)^2}{15} + \frac{(319)^2}{15} + \frac{(309)^2}{15} = 22,375.6250 \\
 \frac{\Sigma B^2}{\Sigma ar_b} &= \frac{B_1^2}{ar_{a1}} + \frac{B_2^2}{ar_{a2}} + \dots + \frac{B_4^2}{ar_{a4}} = \frac{(363)^2}{10} + \frac{(247)^2}{12} + \dots + \frac{(218)^2}{11} = 24,486.6803 \\
 \frac{\Sigma \Sigma (AB)^2}{\Sigma r_{ab}} &= \frac{A_1B_1^2}{r_{a1b1}} + \frac{A_1B_2^2}{r_{a1b2}} + \dots + \frac{A_3B_4^2}{r_{a3b4}} = \frac{(73)^2}{4} + \frac{(100)^2}{4} + \dots + \frac{(76)^2}{4} = 22,833.2500 \\
 \Sigma \Sigma \Sigma (ABR)^2 &= (A_1B_1R_1)^2 + (A_1B_1R_2)^2 + \dots + (A_3B_4R_4)^2 = (26^2 + 29^2 + \dots + 29^2) = 23,374.0000 \\
 \text{C.F.} &= \frac{GT^2}{\Sigma ab_r} = \frac{(1000)^2}{(3 + 4 + \dots + 4)} = 22,222.2222
 \end{aligned}$$

Sum of Squares

$$\text{ToSS} = \Sigma \Sigma \Sigma (ABR)^2 - \text{C.F.} = 23,374.0000 - 22,222.2222 = 1,151.7778$$

$$\text{ASS} = \Sigma \frac{A^2}{\Sigma br_a} - \text{C.F.} = 22,375.6250 - 22,222.2222 = 152.8444$$

$$BSS = \sum \frac{B^2}{\sum ar_b} - C.F. = 24,486.6803 - 22,222.2222 = \boxed{284.4581}$$

$$ABSS = \sum \sum \frac{(AB)^2}{\sum r_{ab}} - \sum \frac{A^2}{\sum br_a} - \sum \frac{B^2}{\sum ar_b} + C.F.$$

$$= 22,833.2500 - 22,375.0667 - 22,486.6803 + 22,222.2222 = \boxed{193.7253}$$

$$ESS = \sum \sum \sum (ABR)^2 - \sum \sum \frac{(AB)^2}{\sum r_{ab}} = 23,374.0000 - 22,833.2500 = \boxed{540.7500}$$

Mean Squares

$$\text{Factor A Mean Square (AMS)} = \frac{ASS}{Adf} = \frac{152.8444}{2} = \boxed{76.4222}$$

$$\text{Factor B Mean Square (BMS)} = \frac{BSS}{Bdf} = \frac{264.4581}{3} = \boxed{88.1527}$$

$$\text{A x B Mean Square (ABMS)} = \frac{ABSS}{ABdf} = \frac{193.7253}{6} = \boxed{32.2875}$$

$$\text{Error Mean Square (EMS)} = \frac{ESS}{Edf} = \frac{540.7500}{33} = \boxed{15.9004}$$

F-computed

$$\text{Factor A F-computed (A Fc)} = \frac{AMS}{EMS} = \frac{76.4222}{15.9004} = \boxed{4.81}$$

$$\text{Factor B F-computed (B Fc)} = \frac{BMS}{EMS} = \frac{88.1527}{15.9004} = \boxed{5.44}$$

$$\text{A x B F-computed (A x B Fc)} = \frac{ABMS}{EMS} = \frac{32.2875}{15.9004} = \boxed{2.03}$$

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Factor A	2	152.8444	76.4222	4.81 *	5.25	3.26
Factor B	3	264.4581	88.1527	5.54 **	4.38	2.86
A x B	6	193.7253	32.2875	2.03 ns	3.35	2.36
Error	33	540.7500	15.9004			
TOTAL	44	1,151.7778				

$$C.V. = \frac{\sqrt{EMS}}{\text{Mean}} \times 100 = \frac{\sqrt{15.9004}}{22.2} \times 100 = \boxed{17.9\%}$$

Mean Comparison

A x B SUMMARY TABLE

	B₁	B₂	B₃	B₄	A MEANS
A₁	24.3	25.0	28.0	21.9	24.8
A₂	26.3	19.5	20.3	18.3	21.3
A₃	28.3	17.3	19.8	18.3	20.6
B MEANS	26.3	20.6	22.7	19.8	

Comparison using LSD

Because of the difference in the number of observations to compute any particular mean, the LSD formula may have different divisors depending on which 2 means are to be compared. As such, the comparisons to be made should be determined first and the specific LSD value is computed per comparison.

Factor A levels

A close examination of the individual observations in the raw data showed that A₁, A₂, & A₃ all have 15 observations each. Assuming that for Factor A, A₁ will be compared with A₂ and with A₃, both pair comparisons have the same LSD formula. Since the F-test showed significant differences among the means, only LSD at 5% probability level will be used.

$$s_d = \sqrt{\frac{2MSE}{\frac{n_1 + n_2}{2}}} = \sqrt{\frac{2 \times 15.5694}{\frac{15 + 15}{2}}} = \boxed{1.441}$$

where **n₁** = number of observations for Mean 1; **n₂** = number of observations for Mean 2

$$LSD_{0.05, 33 \text{ df}} = 2.060 \times 1.395 = \boxed{2.93}$$

Mean

A1 = 24.8

A2 = 21.3 *

A3 = 20.6 *

Since the difference between A₁ and A₂ as well as A₁ and A₃ are bigger than LSD 5%, A₁ is declared significantly different from either A₂ or A₃.

Factor B levels

A close examination of the individual observations in the raw data showed that B₁ has 9 observations, B₂ & B₃ have 12, and B₄ has 11. Since the F-test showed highly significant differences among the means, LSD at 1% or 5% probability level may be used. Assuming that for Factor B, B₁ will be compared with B₂, B₃ and B₄, each pair comparison will have different LSD formula. To compare B₁ (n = 9) with B₂ (n = 12):

$$s_d = \sqrt{\frac{2MSE}{\frac{n_1 + n_2}{2}}} = \sqrt{\frac{2 \times 15.9004}{\frac{9 + 12}{2}}} = \boxed{1.740}$$

$$\text{LSD}_{0.05, 33 \text{ df}} = 2.031 \times 1.740 = \boxed{3.53}$$

$$\text{LSD}_{0.01, 33 \text{ df}} = 2.736 \times 1.740 = \boxed{4.76}$$

To compare B_1 ($n = 9$) with B_4 ($n = 11$):

$$s_d = \sqrt{\frac{\frac{2\text{MSE}}{n_1 + n_2}}{2}} = \sqrt{\frac{2 \times 15.9004}{9 + 11}} = \boxed{1.783}$$

$$\text{LSD}_{0.05, 33 \text{ df}} = 2.031 \times 1.783 = \boxed{3.62}$$

$$\text{LSD}_{0.01, 33 \text{ df}} = 2.736 \times 1.783 = \boxed{4.88}$$

DMRT may be used to compare mean differences. The pooled Error Mean Square will be used to compute the desired R_p values, however, these values are estimates due to differences in number of observations per mean.

To compare the 3 A treatment means:

$$s_d = \mathbf{1.441}$$

Error df = 33

P	r_p	R_p
2	2.88	2.93
3	3.03	3.09

Comparing the 3 means using the appropriate R_p value:

$$A_1 = 25.1 \text{ a}$$

$$A_3 = 21.2 \text{ b}$$

$$A_2 = 21.0 \text{ b}$$

To compare the 4 B treatment means:

$$s_d = \mathbf{1.611}$$

Error df = 36

P	r_p	R_p
2	2.87	3.27
3	3.02	3.44
4	3.11	3.54

Comparing the 4 means using the appropriate R_p value:

$$B_1 = 26.8 \text{ a}$$

$$B_4 = 22.4 \text{ b}$$

$$B_2 = 20.6 \text{ b}$$

$$B_3 = 19.6 \text{ b}$$

FACTORIAL IN RANDOMIZED COMPLETE BLOCK DESIGN

Factorial in RCBD is normally used in field experiment involving two factors that are equally important together with their interaction (A x B). It is not an ideal design for treatments with extensive border effect, e.g. spray, watering or chemicals that may be carried easily by wind.

Randomization

The randomization procedure in factorial in RCBD is very similar to RCBD if the factor combination will be considered as single treatment. The steps in randomization are as follows:

1. Determine the number of factor combinations per replication. This is the product of the number of Factor A and number of Factor B
2. Assign a plot number to each factor combination consecutively.
3. Assign the factor combinations to the experimental plots using a randomization scheme.
 - A. Using the table of random numbers
 1. Before randomization, decide which way will the numbers be assigned to the plots: horizontal or vertical. With horizontal assignment, will it be left to right or right to left? With vertical assignment, will it be top to bottom or bottom to top? With eyes closed, point the finger to the table of random numbers.
 2. Copy the middle 3 digits of all the consecutive 5-digit numbers corresponding to the number of factor combinations per block.
 3. Rank the 3-digit numbers from 1 to the highest number of factor combinations.
 4. Arrange the 3-digit numbers consecutively from lowest to highest but be sure to carry the assigned rank. The sequence of the ranks will now serve as the randomized numbers.
 5. Assign the ranks corresponding to the number of factor combinations of the first block. The first block may look like this:

BLK 1	A ₂ B ₄	A ₃ B ₁	A ₂ B ₁	A ₁ B ₃					BLK 2
	A ₂ B ₁	A ₂ B ₂	A ₃ B ₂	A ₃ B ₃					
	A ₁ B ₂	A ₁ B ₄	A ₁ B ₁	A ₃ B ₄					
BLK 3									BLK 4

6. Repeat steps 1–5 for the other blocks.
- B. Using draw lots.
 1. Prepare pieces of papers corresponding to the number of factor combinations.
 2. Write the factor combinations per block in each of the paper.
 3. Mix the papers thoroughly in a container.
 4. Without looking in the box, draw a paper and assign the factor combination name on the first experimental unit of the first block.
 5. Without returning the paper previously drawn, continue drawing individual papers

and assign the corresponding factor combination name until all the factor combinations have been assigned to the first block.

6. Repeat steps 1–5 for the remaining blocks.

As a general rule, try to make each block as square as possible to reduce the distance between the two farthest plots. In a very long block, higher soil variability may occur between two plots on both ends of the area. Using the randomization discussed above, a possible lay-out of a 3 x 4 factorial with 4 replications is as follows if the plots are short and wide, or the soil fertility gradient is not known:

BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4	A ₃ B ₂	A ₂ B ₃	A ₁ B ₃	A ₃ B ₃
				A ₂ B ₁	A ₁ B ₂	A ₃ B ₄	A ₁ B ₄
				A ₁ B ₁	A ₁ B ₄	A ₂ B ₁	A ₂ B ₄
				A ₁ B ₂	A ₃ B ₃	A ₂ B ₃	A ₂ B ₂
BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4	A ₃ B ₁	A ₂ B ₄	A ₃ B ₂	A ₃ B ₁
				A ₂ B ₂	A ₃ B ₄	A ₁ B ₁	A ₁ B ₂
				A ₁ B ₃	A ₃ B ₃	A ₃ B ₄	A ₃ B ₂
				A ₂ B ₁	A ₃ B ₂	A ₂ B ₃	A ₂ B ₁
BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4	A ₃ B ₁	A ₂ B ₂	A ₁ B ₂	A ₁ B ₄
				A ₂ B ₄	A ₂ B ₁	A ₃ B ₃	A ₃ B ₁
				A ₁ B ₄	A ₃ B ₄	A ₂ B ₂	A ₂ B ₄
				A ₁ B ₂	A ₁ B ₁	A ₁ B ₁	A ₁ B ₃

Or it could be this way:

A ₂ B ₄	A ₃ B ₁	A ₂ B ₁	A ₁ B ₃	A ₂ B ₂	A ₃ B ₁	A ₁ B ₂	A ₁ B ₁
A ₂ B ₁	A ₂ B ₂	A ₃ B ₂	A ₃ B ₃	A ₃ B ₄	A ₂ B ₄	A ₃ B ₃	A ₁ B ₄
A ₁ B ₂	A ₁ B ₄	A ₁ B ₁	A ₃ B ₄	A ₁ B ₂	A ₂ B ₃	A ₂ B ₁	A ₃ B ₂
A ₂ B ₃	A ₃ B ₄	A ₂ B ₁	A ₃ B ₂	A ₁ B ₄	A ₃ B ₃	A ₃ B ₄	A ₁ B ₃
A ₁ B ₁	A ₂ B ₂	A ₃ B ₃	A ₁ B ₂	A ₁ B ₁	A ₃ B ₂	A ₂ B ₃	A ₂ B ₁
A ₁ B ₃	A ₂ B ₄	A ₃ B ₁	A ₁ B ₄	A ₁ B ₂	A ₃ B ₁	A ₂ B ₂	A ₂ B ₄

If the plots are long and narrow, or there is one-directional soil fertility gradient, or a the wind direction will have an effect of the treatment, the following lay-out may be applied with the length of the plot perpendicular to the strong gradient:

BLOCK 4	A_3B_3	A_2B_3	A_3B_2	A_3B_3
	A_3B_2	A_1B_2	A_2B_1	A_1B_4
	A_2B_2	A_1B_4	A_3B_1	A_2B_4
	A_2B_1	A_3B_3	A_2B_4	A_1B_3
BLOCK 3	A_3B_4	A_2B_4	A_3B_4	A_1B_3
	A_1B_1	A_3B_4	A_2B_3	A_1B_3
	A_1B_3	A_3B_2	A_2B_1	A_1B_1
	A_2B_1	A_2B_1	A_1B_1	A_1B_2
BLOCK 2	A_3B_1	A_2B_2	A_3B_3	A_2B_2
	A_3B_2	A_3B_1	A_1B_2	A_2B_2
	A_2B_4	A_1B_4	A_3B_3	A_2B_2
	A_1B_1	A_2B_2	A_3B_3	A_2B_2
BLOCK 1	A_1B_1	A_3B_2	A_2B_3	A_2B_1
	A_3B_3	A_1B_4	A_2B_3	A_2B_1
	A_2B_4	A_1B_4	A_2B_3	A_2B_1
	A_3B_3	A_1B_4	A_2B_3	A_2B_1

Consider an experiment based on the lay-out above involving 3 levels of Factor A, 4 levels of Factor B, and 4 replications. The hypothetical data are in the next page.

Basic values

$$\begin{aligned}
 a &= \text{number or levels of Factor A} & &= 3 \\
 b &= \text{number or levels of Factor B} & &= 4 \\
 r &= \text{number of replications} & &= 4 \\
 GT &= (A_1B_1R_1) + (A_1B_1R_2) + \dots + (A_3B_4R_4) = 26 + 29 + \dots + 29 & &= 1076 \\
 \bar{X} &= \frac{(GT)}{abr} = \frac{1,076}{(3 \times 4 \times 4)} & &= 22.4
 \end{aligned}$$

Degrees of Freedom

$$\begin{aligned}
 \text{Block} &= (r - 1) = (4 - 1) & &= 3 \\
 \text{Factor A} &= (a - 1) = (3 - 1) & &= 2 \\
 \text{Factor B} &= (b - 1) = (4 - 1) & &= 3 \\
 A \times B &= (a - 1)(b - 1) = (3 - 1)(4 - 1) & &= 6 \\
 \text{Error} &= (ab - 1)(r - 1) = (12 - 1)(4 - 1) & &= 33 \\
 \text{Total} &= (abr - 1) = 3 \times 4 \times 4 - 1 & &= 47
 \end{aligned}$$

FACTOR A	FACTOR B	REP I	REP II	REP III	REP IV	TOTAL	MEAN
A ₁	B ₁	28	29	23	22	102	25.5
A ₁	B ₂	25	24	27	24	100	25.0
A ₁	B ₃	26	28	31	27	112	28.0
A ₁	B ₄	23	24	17	23	87	21.8
A ₂	B ₁	27	31	20	27	105	26.2
A ₂	B ₂	16	20	22	20	78	19.5
A ₂	B ₃	17	23	16	25	81	20.2
A ₂	B ₄	19	18	17	18	72	18.0
A ₃	B ₁	30	29	23	33	115	28.8
A ₃	B ₂	16	14	16	23	69	17.2
A ₃	B ₃	18	16	20	25	79	19.8
A ₃	B ₄	17	20	10	29	76	19.0
TOTAL		262	276	242	296	1,076	22.4

A x B SUMMARY TABLE

	B ₁	B ₂	B ₃	B ₄	A TOTALS	A MEANS
A ₁	102	100	112	87	401	25.1
A ₂	105	78	81	72	336	21.0
A ₃	115	69	79	76	339	21.2
B TOTALS	322	247	272	235	1,076	
B MEANS	26.8	20.6	22.7	19.6		

Squares

$$\text{C.F.} = \frac{GT^2}{abr} = \frac{(1076)^2}{(3 \times 4 \times 4)} = \boxed{24,120.3333}$$

$$\frac{\sum R^2}{ab} = \frac{R_1^2 + R_2^2 + \dots + R_4^2}{ab} = \frac{262^2 + 276^2 + \dots + 296^2}{3 \times 4} = \boxed{24,250.0000}$$

$$\frac{\sum A^2}{br} = \frac{A_1^2 + A_2^2 + A_3^2}{br} = \frac{401^2 + 336^2 + 339^2}{4 \times 4} = \boxed{24,288.6250}$$

$$\frac{\sum B^2}{ar} = \frac{B_1^2 + B_2^2 + B_3^2 + B_4^2}{ar} = \frac{322^2 + 247^2 + 272^2 + 235^2}{3 \times 4} = \boxed{24,491.8333}$$

$$\frac{\sum \sum (AB)^2}{r} = \frac{(A_1B_1)^2 + (A_1B_2)^2 + \dots + (A_3B_4)^2}{r} = \frac{102^2 + 100^2 + \dots + 76^2}{3} = \boxed{24,843.5000}$$

$$\sum \sum \sum (ABR)^2 = (A_1B_1R_1)^2 + (A_1B_1R_2)^2 + \dots + (A_3B_4R_4)^2 = 26^2 + 29^2 + \dots + 29^2 = \boxed{25,404.0000}$$

Sum of Squares

$$\text{ToSS} = \sum \sum \sum (\text{ABR})^2 - \text{C.F.} = 25,404.0000 - 24,120.3333 = \boxed{1,283.6667}$$

$$\text{RSS} = \frac{\sum R^2}{ab} - \text{C.F.} = 24,250.0000 - 24,120.3333 = \boxed{129.6667}$$

$$\text{ASS} = \frac{\sum A^2}{br} - \text{C.F.} = 24,288.6250 - 24,120.3333 = \boxed{168.2917}$$

$$\text{BSS} = \frac{\sum B^2}{ar} - \text{C.F.} = 24,491.8333 - 24,120.3333 = \boxed{371.5000}$$

$$\begin{aligned} \text{ABSS} &= \frac{\sum \sum (\text{AB})^2}{r} - \frac{\sum A^2}{br} - \frac{\sum B^2}{ar} + \text{C.F.} \\ &= 24,843.5000 - 24,288.6250 - 24,491.9833 + 24,120.3333 = \boxed{183.3750} \end{aligned}$$

$$\begin{aligned} \text{ESS} &= \sum \sum \sum (\text{ABR})^2 - \frac{\sum \sum (\text{AB})^2}{r} - \frac{\sum R^2}{ab} + \text{C.F.} \\ &= 25,404.0000 - 24,843.5000 - 24,491.8333 + 24,120.3333 = \boxed{430.8333} \end{aligned}$$

Mean Squares

$$\text{Block Mean Square (RMS)} = \frac{\text{RSS}}{\text{R df}} = \frac{129.6667}{3} = \boxed{43.2222}$$

$$\text{Factor A Mean Square (AMS)} = \frac{\text{ASS}}{\text{A df}} = \frac{168.2917}{2} = \boxed{84.1458}$$

$$\text{Factor B Mean Square (BMS)} = \frac{\text{BSS}}{\text{B df}} = \frac{371.5000}{3} = \boxed{123.8333}$$

$$\text{A x B Mean Square (ABMS)} = \frac{\text{ABSS}}{\text{AB df}} = \frac{183.3750}{6} = \boxed{30.5625}$$

$$\text{Error Mean Square (MSE)} = \frac{\text{ESS}}{\text{E df}} = \frac{430.8333}{33} = \boxed{13.0556}$$

F-computed

$$\text{Block F computed (R Fc)} = \frac{\text{RMS}}{\text{MSE}} = \frac{43.2222}{13.0556} = \boxed{3.31}$$

$$\text{Factor A F-computed (A Fc)} = \frac{\text{AMS}}{\text{MSE}} = \frac{84.1458}{13.0556} = \boxed{6.45}$$

$$\text{Factor B F-computed (B Fc)} = \frac{\text{BMS}}{\text{MSE}} = \frac{123.8333}{13.0556} = \boxed{9.49}$$

$$\text{A x B F-computed (A x B Fc)} = \frac{\text{ABMS}}{\text{MSE}} = \frac{30.5625}{13.0556} = \boxed{2.34}$$

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	3	129.6667	43.2222	3.31 *	4.44	2.89
Factor A	2	168.6677	84.1458	6.45 **	5.31	3.29
Factor B	3	371.5000	123.8333	9.49 **	4.44	2.89
A x B	6	183.3750	30.5625	2.34 ns	3.40	2.39
Error	36	430.8333	13.0556			
TOTAL	47	1,283.6667				

$$\text{C.V.} = \frac{\sqrt{\text{MSE}}}{\text{Mean}} \times 100 = \frac{\sqrt{13.0556}}{22.4} \times 100 = \boxed{16.1\%}$$

The results in both factorial in CRD and RCBD showed that highly significant differences among Factors A & B means but no significant differences were noted among the interaction means between A & B. This indicates that the variation in Factor A means may be attributed to the effect of the different levels of that factor and is not significantly affected the different levels of Factor B. At the same time, the variation in Factor B means may be attributed to the effect of the different levels of that factor and is not significantly affected the different levels of Factor A. The coefficient of variation (C.V.) of Factorial in CRD was higher than that of Factorial in RCBD.

Mean Comparison using LSD

Factor A levels

$$S_d = \sqrt{\frac{2\text{MSE}}{b \times r}} = \sqrt{\frac{2 \times 13.0556}{4 \times 4}} = \boxed{1.275}$$

$$\text{LSD}_{0.05, 36 \text{ df}} = 2.031 \times 1.275 = \boxed{2.59}$$

$$\text{LSD}_{0.01, 36 \text{ df}} = 2.727 \times 1.275 = \boxed{3.48}$$

Any of the two LSD values will be used to compare the Treatment Means of the 3 Factor A levels. Since there are 3 treatments, only 2 valid pair comparisons may be made. Comparing A₁ with A₂, and A₁ with A₃ shows that A₁ is significantly different from A₂ and also from A₃ at both 1% and 5% probability level.

Factor B levels

$$s_d = \sqrt{\frac{2\text{MSE}}{a \times r}} = \sqrt{\frac{2 \times 15.5694}{3 \times 4}} = \boxed{1.611}$$

$$\text{LSD}_{0.05, 30 \text{ df}} = 2.060 \times 1.611 = \boxed{2.99}$$

$$\text{LSD}_{0.01, 30 \text{ df}} = 2.750 \times 1.611 = \boxed{4.01}$$

Any of the two LSD values will be used to compare the Treatment Means of the 3 Factor B levels depending on the desired probability level. Since there are 4 treatments, only 3 valid pair comparisons may be made. Comparing B₁ with B₂, B₃ and B₄ shows that at 1% probability level, B₁ is highly significantly different from all 3 treatment means.

Mean Comparison using DMRT

If DMRT is desired, the same procedure as in previous example in RCBD may be used using the corresponding s_d values computed above.

To compare the 3 A treatment means:

$$s_d = 1.275$$

Error df = 36

P	r_p	R_p
2	2.87	2.59
3	3.02	2.72

Comparing the 3 means using the appropriate R_p value:

$$A_1 = 25.1 \text{ a}$$

$$A_3 = 21.2 \text{ b}$$

$$A_2 = 21.0 \text{ b}$$

To compare the 4 B treatment means:

$$s_d = 1.611$$

Error df = 36

P	r_p	R_p
2	2.87	3.27
3	3.02	3.44
4	3.11	3.54

Comparing the 4 means using the appropriate R_p value:

$$B_1 = 26.8 \text{ a}$$

$$B_4 = 22.4 \text{ b}$$

$$B_2 = 20.6 \text{ b}$$

$$B_3 = 19.6 \text{ b}$$

SPLIT-PLOT

Features

The main plot is assigned to either the whole experimental area (in CRD) or in each block (in RCBD) while the subplots are assigned and randomized in each main plot. With the randomization of the subplot in each main plot, the interaction (MP x SP) is also created (therefore, the size of the subplot is always the same as that of the interaction). The size of the subplot (and the interaction) will always be smaller than the main plot. With this design, the subplot and the interaction between the main plot and the subplot have higher degree of precision than the main plot.

Application

Split-plot design is used in experiment involving two factors where one factor is more important than the other such that a more sensitive detection of significant differences is desired in that factor. The factor with less importance is assigned the main plot while the factor with greater importance, the subplot. As an example, an experiment involving different varieties of tomato are being evaluated using different fertilizer level. The researcher is more interested in the difference among varieties than the effect of fertilizer. In this case, a split-plot in RCBD is used with the fertilizer level assigned as main plot and the different varieties are the subplot.

The design is also used for factors that require large plot sizes to avoid high border effect. This design is most applicable for chemical application or irrigation frequency if these factors are of lesser importance than the other factor being evaluated.

SPLIT-PLOT IN COMPLETELY RANDOMIZED DESIGN

Randomization

The basic differences between factorial and split-plot in CRD are in lay-out and randomization. In factorial in CRD, all factor combinations are assigned and randomized in the whole experimental area (since there are no blocks). In split-plot, only the main plots are assigned and randomized in the experimental area. After randomizing all the main plots, the subplot is randomized in each main plot. With this randomization scheme, the size of the main plot is much larger than that of the subplot so it is expected that the degree of precision is less in the main plot than in the subplot. Moreover, the size of the subplot is the same as that of the interaction so the degree of precision of the subplot is equal to that of the interaction. The procedure in randomization is as follows:

1. Determine the total number of Main Plot x Replication combinations in the experiment. In the example, there are 3 x 4 combinations of Main Plot x Replication.
2. Assign a plot number to each main plot treatment consecutively.
3. Assign the main plot treatments to the experimental plots using a randomization scheme as discussed in CRD:
 - A. Table of random numbers
 - B. Draw lots

A possible randomization of Main Plot x Replication is shown below:

R_1A_1	R_3A_3	R_2A_3	R_1A_2
R_2A_1	R_4A_2	R_1A_3	R_4A_3
R_3A_2	R_3A_1	R_2A_2	R_4A_1

4. Determine the number of subplots.
5. In each Main Plot x Replication combination, randomize the subplot levels independent of the other Main Plot x Replication combinations.

$R_1A_1B_1$	R_3A_3	R_2A_3	R_1A_2
$R_1A_1B_3$			
$R_1A_1B_2$			
$R_1A_1B_4$			
R_2A_1	R_4A_2	R_1A_3	R_4A_3
R_3A_2	R_3A_1	R_2A_2	R_4A_1

6. Repeat the procedure in the other combinations. When completed, a possible lay-out of a 3 x 4 split-plot in CRD with 4 replications is as follows:

$R_1A_1B_1$	$R_3A_3B_4$	$R_2A_3B_2$	$R_1A_2B_3$
$R_1A_1B_3$	$R_3A_3B_2$	$R_2A_3B_3$	$R_1A_2B_2$
$R_1A_1B_2$	$R_3A_3B_1$	$R_2A_3B_4$	$R_1A_2B_4$
$R_1A_1B_4$	$R_3A_3B_3$	$R_2A_3B_1$	$R_1A_2B_1$
$R_2A_1B_4$	$R_4A_2B_1$	$R_1A_3B_2$	$R_4A_3B_2$
$R_2A_1B_1$	$R_4A_2B_3$	$R_1A_3B_4$	$R_4A_3B_1$
$R_2A_1B_2$	$R_4A_2B_2$	$R_1A_3B_3$	$R_4A_3B_4$
$R_2A_1B_3$	$R_4A_2B_4$	$R_1A_3B_1$	$R_4A_3B_3$
$R_3A_2B_2$	$R_3A_1B_4$	$R_2A_2B_1$	$R_4A_1B_3$
$R_3A_2B_3$	$R_3A_1B_1$	$R_2A_2B_4$	$R_4A_1B_4$
$R_3A_2B_4$	$R_3A_1B_2$	$R_2A_2B_3$	$R_4A_1B_1$
$R_3A_2B_1$	$R_3A_1B_3$	$R_2A_2B_2$	$R_4A_1B_2$

The lay-out may also be like these depending on available area:

R ₁ A ₁ B ₁	R ₃ A ₃ B ₄	R ₂ A ₃ B ₂
R ₁ A ₁ B ₂	R ₃ A ₃ B ₂	R ₂ A ₃ B ₃
R ₁ A ₁ B ₃	R ₃ A ₃ B ₁	R ₂ A ₃ B ₄
R ₁ A ₁ B ₄	R ₃ A ₃ B ₃	R ₂ A ₃ B ₁
R ₂ A ₁ B ₄	R ₄ A ₂ B ₁	R ₁ A ₃ B ₂
R ₂ A ₁ B ₁	R ₄ A ₂ B ₃	R ₁ A ₃ B ₄
R ₂ A ₁ B ₂	R ₄ A ₂ B ₂	R ₁ A ₃ B ₃
R ₂ A ₁ B ₃	R ₄ A ₂ B ₄	R ₁ A ₃ B ₁
R ₃ A ₂ B ₂	R ₃ A ₁ B ₄	R ₂ A ₂ B ₁
R ₃ A ₂ B ₃	R ₃ A ₁ B ₁	R ₂ A ₂ B ₄
R ₃ A ₂ B ₄	R ₃ A ₁ B ₂	R ₂ A ₂ B ₃
R ₃ A ₂ B ₁	R ₃ A ₁ B ₃	R ₂ A ₂ B ₂
R ₁ A ₂ B ₃	R ₄ A ₃ B ₂	R ₄ A ₁ B ₃
R ₁ A ₂ B ₂	R ₄ A ₃ B ₁	R ₄ A ₁ B ₄
R ₁ A ₂ B ₄	R ₄ A ₃ B ₄	R ₄ A ₁ B ₁
R ₁ A ₂ B ₁	R ₄ A ₃ B ₃	R ₄ A ₁ B ₂

or

R ₃ A ₃ B ₄	R ₁ A ₂ B ₃
R ₃ A ₃ B ₂	R ₁ A ₂ B ₂
R ₃ A ₃ B ₁	R ₁ A ₂ B ₄
R ₃ A ₃ B ₃	R ₁ A ₂ B ₁
R ₄ A ₂ B ₁	R ₄ A ₃ B ₂
R ₄ A ₂ B ₃	R ₄ A ₃ B ₁
R ₄ A ₂ B ₂	R ₄ A ₃ B ₄
R ₄ A ₂ B ₄	R ₄ A ₃ B ₃
R ₃ A ₁ B ₄	R ₄ A ₁ B ₃
R ₃ A ₁ B ₁	R ₄ A ₁ B ₄
R ₃ A ₁ B ₂	R ₄ A ₁ B ₁
R ₃ A ₁ B ₃	R ₄ A ₁ B ₂
R ₂ A ₃ B ₂	R ₁ A ₁ B ₁
R ₂ A ₃ B ₃	R ₁ A ₁ B ₂
R ₂ A ₃ B ₄	R ₁ A ₁ B ₃
R ₂ A ₃ B ₁	R ₁ A ₁ B ₄
R ₁ A ₃ B ₂	R ₂ A ₁ B ₄
R ₁ A ₃ B ₄	R ₂ A ₁ B ₁
R ₁ A ₃ B ₃	R ₂ A ₁ B ₂
R ₁ A ₃ B ₁	R ₂ A ₁ B ₃
R ₂ A ₂ B ₁	R ₃ A ₂ B ₂
R ₂ A ₂ B ₄	R ₃ A ₂ B ₃
R ₂ A ₂ B ₃	R ₃ A ₂ B ₄
R ₂ A ₂ B ₂	R ₃ A ₂ B ₁

Since the main plot (Factor A) is randomized in the whole experimental area and each subplot is randomized in each main plot, the error used to test the main plot should be different from the error to test the subplot (and interaction). Therefore, there should be two (2) error terms in split-plot.

The same set of data as in Factorial will be used in split-plot to have a comparison of the two designs. The data are as follows:

MAIN PLOT (A)	SUBPLOT (B)	REP I	REP II	REP III	REP IV	TOTAL	MEAN
A ₁	B ₁	28	29	23	22	102	25.5
A ₁	B ₂	25	24	27	24	100	25.0
A ₁	B ₃	26	28	31	27	112	28.0
A ₁	B ₄	23	24	17	23	87	21.8
A ₂	B ₁	27	31	20	27	105	26.2
A ₂	B ₂	16	20	22	20	78	19.5
A ₂	B ₃	17	23	16	25	81	20.2
A ₂	B ₄	19	18	17	18	72	18.0
A ₃	B ₁	30	29	23	33	115	28.8
A ₃	B ₂	16	14	16	23	69	17.2
A ₃	B ₃	18	16	20	25	79	19.8
A ₃	B ₄	17	20	10	29	76	19.0
TOTAL						1,076	22.4

Aside from A x B summary table, A x R summary table is needed in Split-plot.

A x B SUMMARY TABLE

	B ₁	B ₂	B ₃	B ₄	A TOTALS	A MEANS
A ₁	102	100	112	87	401	25.1
A ₂	105	78	81	72	336	21.0
A ₃	115	69	79	76	339	21.2
B TOTALS	322	247	272	235	1,076	
B MEANS	26.8	20.6	22.7	19.6		

A x R SUMMARY TABLE

	R ₁	R ₂	R ₃	R ₄	A TOTALS	A MEANS
A ₁	102	105	98	96	401	25.1
A ₂	79	92	75	90	336	21.0
A ₃	81	79	69	110	339	21.2
R TOTALS	262	276	242	296	1,076	

Basic values

$$\begin{aligned}
 a &= \text{number or levels of Factor A} & &= 3 \\
 b &= \text{number or levels of Factor B} & &= 4 \\
 r &= \text{number of replications} & &= 4 \\
 GT &= (A_1B_1R_1) + (A_1B_1R_2) + \dots + (A_3B_4R_4) = 26 + 29 + \dots + 29 & &= 1,076 \\
 \bar{X} &= \frac{(GT)}{abr} = \frac{1,076}{(3 \times 4 \times 4)} & &= 22.4
 \end{aligned}$$

Degrees of Freedom

$$\begin{aligned}
 \text{Main-plot A} &= (a - 1) = (3 - 1) & &= 2 \\
 \text{Error (a)} &= a(r - 1) = 3(4 - 1) & &= 9 \\
 \text{Subplot B} &= (b - 1) = (4 - 1) & &= 3 \\
 \text{A x B} &= (a - 1)(b - 1) = (3 - 1)(4 - 1) & &= 6 \\
 \text{Error (b)} &= a(b - 1)(r - 1) = 3(4 - 1)(4 - 1) & &= 27 \\
 \text{Total} &= (abr - 1) = 3 \times 4 \times 4 - 1 & &= 47
 \end{aligned}$$

Error (a) will be used to test the significance of Factor A while Error (b) will be used to test significance of Factor B and A x B.

Squares

$$\text{C.F.} = \frac{GT^2}{abr} = \frac{1076^2}{3 \times 4 \times 4} = \boxed{24,120.3333}$$

$$\frac{\Sigma A^2}{br} = \frac{A_1^2 + A_2^2 + A_3^2}{br} = \frac{401^2 + 336^2 + 330^2}{4 \times 4} = \boxed{24,266.6250}$$

$$\frac{\Sigma B^2}{ar} = \frac{B_1^2 + B_2^2 + B_3^2 + B_4^2}{ar} = \frac{322^2 + 247^2 + 272^2 + 235^2}{3 \times 4} = \boxed{24,491.8333}$$

$$\frac{\Sigma \Sigma (AB)^2}{r} = \frac{(A_1B_1)^2 + (A_1B_2)^2 + \dots + (A_3B_4)^2}{r} = \frac{102^2 + 100^2 + \dots + 76^2}{3} = \boxed{24,843.5000}$$

$$\frac{\Sigma \Sigma (AR)^2}{b} = \frac{(A_1R_1)^2 + (A_1R_2)^2 + \dots + (A_3R_4)^2}{b} = \frac{102^2 + 105^2 + \dots + 110^2}{4} = \boxed{24,585.5000}$$

$$\Sigma \Sigma \Sigma (ABR)^2 = (A_1B_1R_1)^2 + (A_1B_1R_2)^2 + \dots + (A_3B_4R_4)^2 = 26^2 + 29^2 + \dots + 29^2 = \boxed{25,404.0000}$$

Sum of Squares

$$\text{ToSS} = \Sigma \Sigma \Sigma (ABR)^2 - \text{C.F.} = 25,404.0000 - 24,120.3333 = \boxed{1,283.6667}$$

$$\text{ASS} = \frac{\Sigma A^2}{br} - \text{C.F.} = 24,266.6250 - 24,120.3333 = \boxed{146.2917}$$

$$\text{E(a) SS} = \frac{\Sigma \Sigma (AR)^2}{b} - \frac{\Sigma A^2}{br} = 24,585.5000 - 24,266.6250 = \boxed{318.8750}$$

$$\text{BSS} = \frac{\Sigma B^2}{ar} - \text{C.F.} = 24,491.8333 - 24,120.3333 = \boxed{371.5000}$$

$$\begin{aligned} \text{ABSS} &= \frac{\Sigma \Sigma (AB)^2}{r} - \frac{\Sigma A^2}{br} - \frac{\Sigma B^2}{ar} + \text{C.F.} \\ &= 24,843.5000 - 24,266.6250 - 24,491.8333 + 24,120.3333 = \boxed{184.3750} \end{aligned}$$

$$\begin{aligned} \text{E(b) SS} &= \Sigma \Sigma \Sigma (ABR)^2 - \frac{\Sigma \Sigma (AB)^2}{r} - \frac{\Sigma \Sigma (AR)^2}{b} + \frac{\Sigma A^2}{br} \\ &= 25,404.0000 - 24,843.5000 - 24,585.5000 + 24,266.6250 = \boxed{263.6250} \end{aligned}$$

Mean Squares

$$\text{Factor A Mean Square (AMS)} = \frac{\text{ASS}}{\text{A df}} = \frac{168.2917}{2} = \boxed{84.1458}$$

$$\text{Error (a) Mean Square (MSE)} = \frac{\text{EaSS}}{\text{Ea df}} = \frac{296.8750}{9} = \boxed{32.9861}$$

$$\text{Factor B Mean Square (BMS)} = \frac{\text{BSS}}{\text{B df}} = \frac{371.5000}{3} = \boxed{123.8333}$$

$$\text{A x B Mean Square (ABMS)} = \frac{\text{ABSS}}{\text{AB df}} = \frac{183.3750}{6} = \boxed{30.5625}$$

$$\text{Error (b) Mean Square (MSE)} = \frac{\text{E(b)SS}}{\text{(b) df}} = \frac{263.6250}{27} = \boxed{9.7639}$$

F-Computed

$$\text{Factor A F-computed (A Fc)} = \frac{\text{AMS}}{\text{E(a)MS}} = \frac{84.1458}{32.9861} = \boxed{2.55}$$

$$\text{Factor B F-computed (B Fc)} = \frac{\text{BMS}}{\text{E(b)MS}} = \frac{123.8333}{9.7639} = \boxed{12.88}$$

$$\text{A x B F-computed (A x B Fc)} = \frac{\text{ABMS}}{\text{E(ab)MS}} = \frac{30.5625}{9.7639} = \boxed{3.13}$$

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Main plot (A)	2	166.2917	84.1458	2.55 ns	8.62	4.26
Error (a)	9	296.8750	32.9861			
Subplot (B)	3	371.5000	123.8333	12.68 **	4.50	2.96
A x B	6	183.3750	30.5625	3.13 *	3.56	2.46
Error (b)	27	263.6250	9.7639			
TOTAL	47	1,283.6667				

Coefficient of Variation (C.V.) may be computed in two ways: pooled C.V. or individual C.V. (one for A and one for B). Pooled (or combined) C.V. is computed by computing the average MSE. Average MSE may not be computed using arithmetic mean because the divisors (Error df) have different values. Weighted mean is therefore used:

$$\text{Pooled MSE} = \frac{296.8750 + 263.6250}{9 + 27} = \boxed{15.5694}$$

$$\text{Pooled C.V.} = \frac{\sqrt{\text{pooled MSE}}}{\text{Mean}} \times 100 = \frac{\sqrt{15.5694}}{22.4} \times 100 = \boxed{17.6\%}$$

Two individual C.V. may be computed as:

$$\text{C.V. (a)} = \frac{\sqrt{\text{MSE(a)}}}{\text{Mean}} \times 100 = \frac{\sqrt{32.9861}}{22.4} \times 100 = \boxed{25.6\%}$$

$$\text{C.V. (b)} = \frac{\sqrt{\text{MSE(b)}}}{\text{Mean}} \times 100 = \frac{\sqrt{9.7639}}{22.4} \times 100 = \boxed{13.9\%}$$

Mean Comparison using LSD

There are 4 possible kinds of pair comparisons in Split-plot:

1. Comparison between two main-plot treatment means. Since main-plot levels were detected to be not significant, there is no need to compute for its LSD. In case they are significantly different, the formula for standard error of mean difference (sd) is:

$$s_d = \sqrt{\frac{2 \times \text{MSE(a)}}{b \times r}}$$

and LSD is computed by multiplying the s_d value with t value at 5% probability level and at Error(a) df

2. Comparison between two subplot treatment means averaged over all main-plot treatments. The same procedure as in LSD for Factorial CRD is applicable. When computed, the values are:

$$s_d = \sqrt{\frac{2 \times \text{MSE(b)}}{a \times r}} = \sqrt{\frac{2 \times 15.5694}{3 \times 4}} = \boxed{1.28}$$

$$\text{LSD}_{0.05, 27 \text{ df}} = 2.031 \times 1.28 = \boxed{2.62}$$

$$\text{LSD}_{0.01, 27 \text{ df}} = 2.727 \times 1.28 = \boxed{3.53}$$

Any of the two LSD values may be used to compare the Treatment Means of the 3 Factor B levels depending on the desired probability level. Since there are 4 treatments, only 3 valid pair comparisons may be made. Comparing B_1 with B_2 , B_3 and B_4 shows that at 1% probability level, B_1 is highly significantly different from all 3 treatment means.

3. Comparison between two subplot treatment means at the same main-plot treatment. When the interaction between main-plot and subplot is detected to be significant or highly significant, differences among the main plot (A) and subplot (B) treatment means are ignored and the subplot treatment means at each main-plot level are compared. In the example, comparison will be made on the 4 B levels at A_1 , the 4 B levels at A_2 , and the 4 B levels at A_3 . The formula is:

$$s_d = \sqrt{\frac{2 \times \text{MSE(b)}}{r}} = \sqrt{\frac{2 \times 15.5694}{4}} = \boxed{2.79}$$

$$\text{LSD}_{0.05, 27 \text{ df}} = 2.031 \times 2.79 = \boxed{5.67}$$

$$\text{LSD}_{0.01, 27 \text{ df}} = 2.727 \times 2.79 = \boxed{7.61}$$

4. Two main-plot means at the same or different subplot treatments. The test still follows the rule that only $(t - 1)$ number of pair comparisons can be made. In the example with 3 main-plot levels, only 2 valid pair comparisons can be made. The formula is:

$$s_d = \sqrt{\frac{2 [(b - 1)MSE(b) + MSE(a)]}{rb}} = \sqrt{\frac{2 [(4 - 1) \times 15.5694 + 32.9861]}{4 \times 4}} = \boxed{2.81}$$

$$LSD_{0.05, 9 \text{ df}} = 2.262 \times 2.81 = \boxed{6.36}$$

$$LSD_{0.01, 9 \text{ df}} = 3.250 \times 2.81 = \boxed{9.13}$$

Mean Comparison using DMRT

A x B Mean Summary Table

	B ₁	B ₂	B ₃	B ₄	A Means
A ₁	25.5	25.0	28.0	21.9	25.1
A ₂	26.3	19.5	20.3	18.0	21.0
A ₃	28.8	17.3	19.8	19.0	21.2
B Means	26.8	20.6	22.7	19.6	

1. Comparison among main-plot treatment means. Since main-plot levels were detected to be not significant, there is no need to compute for its DMRT.
2. Comparison among subplot treatment means averaged over all main-plot treatments. The same procedure as in previous example in Factorial CRD is applicable using the appropriate s_d value. When computed, the R_p values for comparison of subplot treatment means will be:

R_p

$$2 = 2.63$$

$$3 = 2.76$$

$$4 = 2.84$$

And the mean comparison will be:

$$B_1 = 26.8 \text{ a}$$

$$B_4 = 22.7 \text{ b}$$

$$B_2 = 20.6 \text{ bc}$$

$$B_3 = 19.6 \text{ c}$$

3. Comparison among subplot treatment means at the same main-plot treatment. Similar to the procedure in LSD, comparison is done among the 4 B levels at A₁, 4 B levels at A₂, and 4 B levels at A₃.

To illustrate the comparison of B₁, B₂, B₃ & B₄ at A₁ using the computed s_d value of 2.79, the computed R_p values are:

$$2 = 5.73$$

$$3 = 6.02$$

$$4 = 6.18$$

And the mean comparison will be

$$B_3 = 28.0 \text{ a}$$

$$B_1 = 25.5 \text{ ab}$$

$$B_2 = 25.0 \text{ ab}$$

$$B_4 = 21.9 \text{ b}$$

The same procedure may be applied to compare B_1 - B_4 at A_2 and at A_3 .

4. Comparison among main-plot means at the same subplot treatments. Comparison may be done separately among 3 levels of A at B_1 , at B_2 , at B_3 and at B_4 . Using the s_d value derived from formula in Item 4 for LSD computation (2.84), the R_p values can be computed.

$$2 = 5.83$$

$$3 = 6.12$$

To compare A_1 , A_2 & A_3 at B_1 :

$$A_1 = 25.5 \text{ a}$$

$$A_2 = 26.3 \text{ a}$$

$$A_3 = 28.8 \text{ a}$$

To compare A_1 , A_2 & A_3 at B_3 :

$$A_1 = 28.0 \text{ a}$$

$$A_2 = 20.3 \text{ b}$$

$$A_3 = 19.8 \text{ b}$$

SPLIT-PLOT IN RANDOMIZED COMPLETE BLOCK DESIGN

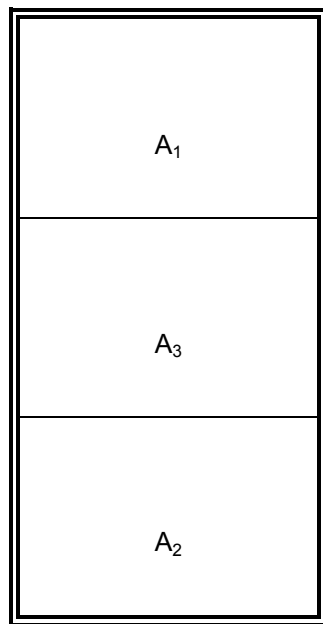
Application

Split-plot in Randomized Complete Block Design is used for field experiment involving two factors tested at the same time with one factor (and its interaction) considered more important than the other factor. It is also recommended if application of treatment of a factor requires a large area to minimize border effect, or to facilitate better management.

Randomization

One of the basic differences between factorial and split-plot in RCBD is in randomization. In factorial in RCBD, randomization of all factor combinations is done in each replication. In split-plot in RCBD, only the main plot is independently randomized in each replication. After randomizing all the main plots, the subplot is randomized in each main plot in each replication. With this randomization scheme, the size of the main plot is much larger than that of the subplot so it is expected that the degree of precision is less in the main plot than in the subplot. Moreover, the size of the subplot is the same as that of the interaction so the degree of precision of the subplot is equal to that of the interaction. The procedure in randomization is as follows:

1. Determine the number of treatments in the main plot. In the example, there are 3 levels in the main plot.
2. In replication 1, assign a number to each main plot treatment consecutively.
3. Assign the main plot treatment to the experimental plots using a randomization scheme as discussed in RCBD:
 - A. Table of random numbers
 - B. Draw lots.



BLOCK 1

4. Determine the number of subplots. In the example, there are 4 subplot levels.
5. Randomize the subplots in each main plot treatment of independent of the other main plot treatments using a randomization scheme discussed in RCBD.

A_1B_1
A_1B_2
A_1B_3
A_1B_4
A_3B_4
A_3B_1
A_3B_2
A_3B_3
A_2B_2
A_2B_3
A_2B_4
A_2B_1

BLOCK 1

6. After finishing randomization of the main plots within Replication 1 and all subplots within each main plots in Replication 1, proceed to Replication 2 and follow Steps 2 to 4.
7. Continue until all replications are completed.

Using the randomization discussed above, a possible lay-out of a 3 x 4 split-plot with 4 replications is as follows.

1. If the plots are long and narrow or if there is a one-way fertility gradient in the area, the following lay-out may be used:

A_1B_1	A_3B_4	A_3B_2	A_2B_3
A_1B_2	A_3B_2	A_3B_3	A_2B_2
A_1B_3	A_3B_1	A_3B_4	A_2B_4
A_1B_4	A_3B_3	A_3B_1	A_2B_1
A_3B_4	A_2B_1	A_1B_2	A_3B_4
A_3B_1	A_2B_3	A_1B_4	A_3B_1
A_3B_2	A_2B_2	A_1B_3	A_3B_2
A_3B_3	A_2B_4	A_1B_1	A_3B_3
A_2B_2	A_1B_4	A_2B_1	A_1B_3
A_2B_3	A_1B_1	A_2B_4	A_1B_4
A_2B_4	A_1B_2	A_2B_3	A_1B_1
A_2B_1	A_1B_3	A_2B_2	A_1B_2
BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4

2. If the fertility of the area is not known, the following lay-out may be used:

A_1B_1	A_3B_4	A_2B_2	A_2B_3	A_1B_4	A_3B_4
A_1B_2	A_3B_2	A_2B_3	A_2B_2	A_1B_1	A_3B_1
A_1B_3	A_3B_1	A_2B_4	A_2B_4	A_1B_2	A_3B_2
A_1B_4	A_3B_3	A_2B_1	A_2B_1	A_1B_3	A_3B_3
A_3B_4	A_2B_1	A_1B_2	A_3B_2	A_2B_1	A_1B_4
A_3B_1	A_2B_3	A_1B_4	A_3B_1	A_2B_4	A_1B_1
A_3B_2	A_2B_2	A_1B_3	A_3B_4	A_2B_3	A_1B_2
A_3B_3	A_2B_4	A_1B_1	A_3B_3	A_2B_2	A_1B_3
BLOCK 3					BLOCK 4
BLOCK 1					BLOCK 2

3. The lay-out may also be like this depending on the dimension and availability of the area:

BLOCK 4	A_1B_1	A_3B_4	A_2B_2
	A_1B_2	A_3B_2	A_2B_3
	A_1B_3	A_3B_1	A_2B_4
	A_1B_4	A_3B_3	A_2B_1
BLOCK 3	A_2B_3	A_1B_4	A_3B_4
	A_2B_2	A_1B_1	A_3B_1
	A_2B_4	A_1B_2	A_3B_2
	A_2B_1	A_1B_3	A_3B_3
BLOCK 2	A_3B_4	A_2B_1	A_1B_2
	A_3B_1	A_2B_3	A_1B_4
	A_3B_2	A_2B_2	A_1B_3
	A_3B_3	A_2B_4	A_1B_1
BLOCK 1	A_3B_2	A_2B_1	A_1B_4
	A_3B_1	A_2B_4	A_1B_1
	A_3B_4	A_2B_3	A_1B_2
	A_3B_3	A_2B_2	A_1B_3

The same set of data used in previous examples will be used to have a comparison between Factorial and Split-Plot in RCBD, and between Split-Plot in CRD and Split-Plot in RCBD.

FACTOR A	FACTOR B	REP I	REP II	REP III	REP IV	TOTAL	MEAN
A ₁	B ₁	28	29	23	22	102	25.5
A ₁	B ₂	25	24	27	24	100	25.0
A ₁	B ₃	26	28	31	27	112	28.0
A ₁	B ₄	23	24	17	23	87	21.8
A ₂	B ₁	27	31	20	27	105	26.2
A ₂	B ₂	16	20	22	20	78	19.5
A ₂	B ₃	17	23	16	25	81	20.2
A ₂	B ₄	19	18	17	18	72	18.0
A ₃	B ₁	30	29	23	33	115	28.8
A ₃	B ₂	16	14	16	23	69	17.2
A ₃	B ₃	18	16	20	25	79	19.8
A ₃	B ₄	17	20	10	29	76	19.0
TOTAL		262	276	242	296	1,076	22.4

A x B SUMMARY TABLE

	B ₁	B ₂	B ₃	B ₄	A TOTALS	A MEANS
A ₁	102	100	112	87	401	25.1
A ₂	105	78	81	72	336	21.0
A ₃	115	69	79	76	339	21.2
B TOTALS	322	247	272	235	1,076	
B MEANS	26.8	20.6	22.7	19.6		

A x R SUMMARY TABLE

	R ₁	R ₂	R ₃	R ₄	A TOTALS	A MEANS
A ₁	102	105	98	96	401	25.1
A ₂	79	92	75	90	336	21.0
A ₃	81	79	69	110	339	21.2
R TOTALS	262	276	242	296	1,076	

Basic values

$$\begin{aligned}
 a &= \text{number or levels of Factor A} & &= 3 \\
 b &= \text{number or levels of Factor B} & &= 4 \\
 r &= \text{number of replications} & &= 4 \\
 \text{GT} &= (A_1B_1R_1) + (A_1B_1R_2) + \dots + (A_3B_4R_4) = 26 + 29 + \dots + 29 & &= 1,076 \\
 \bar{X} &= \frac{(\text{GT})}{abr} = \frac{1,076}{(3 \times 4 \times 4)} & &= 22.4
 \end{aligned}$$

Degrees of Freedom

$$\begin{aligned}
 \text{Block} &= (r - 1) & &= (4 - 1) & &= 3 \\
 \text{Factor A} &= (a - 1) & &= (3 - 1) & &= 2 \\
 \text{Error (a)} &= (a - 1)(r - 1) & &= (3 - 1)(4 - 1) & &= 6
 \end{aligned}$$

Factor B	= (b - 1)	= (4 - 1)	= 3
A x B	= (a - 1)(b - 1)	= (3 - 1)(4 - 1)	= 6
Error (b)	= a(b - 1)(r - 1)	= 3(4 - 1)(4 - 1)	= 27
Total	= (abr - 1)	= 3 x 4 x 4 - 1	= 47

Error (a) will be used to test the significance of Factor A and Block while Error (b) will be used to test the significance of Factor B and A x B.

Squares

$$C.F. = \frac{GT^2}{abr} = \frac{1076^2}{3 \times 4 \times 4} = \boxed{24,120.3333}$$

$$\frac{\Sigma R^2}{ab} = \frac{R_1^2 + R_2^2 + \dots + R_4^2}{ab} = \frac{262^2 + 276^2 + \dots + 296^2}{3 \times 4} = \boxed{24,250.0000}$$

$$\frac{\Sigma A^2}{br} = \frac{A_1^2 + A_2^2 + A_3^2}{br} = \frac{401^2 + 336^2 + 330^2}{4 \times 4} = \boxed{24,266.6250}$$

$$\frac{\Sigma B^2}{ar} = \frac{B_1^2 + B_2^2 + B_3^2 + B_4^2}{ar} = \frac{322^2 + 247^2 + 272^2 + 235^2}{3 \times 4} = \boxed{24,491.8333}$$

$$\frac{\Sigma \Sigma (AB)^2}{r} = \frac{(A_1B_1)^2 + (A_1B_2)^2 + \dots + (A_3B_4)^2}{r} = \frac{102^2 + 100^2 + \dots + 76^2}{3} = \boxed{24,843.5000}$$

$$\frac{\Sigma \Sigma (AR)^2}{b} = \frac{(A_1R_1)^2 + (A_1R_2)^2 + \dots + (A_3R_4)^2}{b} = \frac{102^2 + 105^2 + \dots + 110^2}{4} = \boxed{24,585.5000}$$

$$\Sigma \Sigma \Sigma (ABR)^2 = (A_1B_1R_1)^2 + (A_1B_1R_2)^2 + \dots + (A_3B_4R_4)^2 = 26^2 + 29^2 + \dots + 29^2 = \boxed{25,404.0000}$$

Sum of Squares

$$ToSS = \Sigma \Sigma \Sigma (ABR)^2 - C.F. = 25,404.0000 - 24,120.3333 = \boxed{1,283.6667}$$

$$RSS = \frac{\Sigma R^2}{ab} - C.F. = 24,250.000 - 24,120.3333 = \boxed{129.6667}$$

$$ASS = \frac{\Sigma A^2}{br} - C.F. = 24,266.6250 - 24,120.3333 = \boxed{168.2917}$$

$$E(a) SS = \frac{\Sigma AR^2}{b} - \frac{\Sigma A^2}{br} = 24,585.5000 - 24,266.6250 = \boxed{296.8750}$$

$$BSS = \frac{\Sigma B^2}{ar} - C.F. = 24,491.8333 - 24,120.3333 = \boxed{371.5000}$$

$$ABSS = \frac{\Sigma\Sigma(AB)^2}{r} - \frac{\Sigma A^2}{br} - \frac{\Sigma B^2}{ar} + C.F.$$

$$= 24,843.5000 - 24,288.6250 - 24,491.98333 + 24,120.3333 = \boxed{183.3750}$$

$$E(b)SS = \Sigma\Sigma\Sigma(ABR)^2 - \frac{\Sigma\Sigma(AB)^2}{r} - \frac{\Sigma\Sigma(AR)^2}{b} + \frac{\Sigma(A)^2}{br}$$

$$= 25,404.0000 - 24,843.5000 - 24,585.5000 + 24,288.6250 = \boxed{263.6250}$$

Mean Squares

$$\text{Block Mean Square (RMS)} = \frac{RSS}{R \text{ df}} = \frac{129.6667}{3} = \boxed{43.2222}$$

$$\text{Factor A Mean Square (AMS)} = \frac{ASS}{A \text{ df}} = \frac{168.2917}{2} = \boxed{84.1458}$$

$$\text{Error (a) Mean Square (MSE)} = \frac{E(a)SS}{E(a) \text{ df}} = \frac{167.2083}{6} = \boxed{27.8681}$$

$$\text{Factor B Mean Square (BMS)} = \frac{BSS}{B \text{ df}} = \frac{371.5000}{3} = \boxed{123.8333}$$

$$\text{A x B Mean Square (ABMS)} = \frac{ABSS}{AB \text{ df}} = \frac{183.3750}{6} = \boxed{30.5625}$$

$$\text{Error (b) Mean Square (MSE)} = \frac{E(b)SS}{E(b) \text{ df}} = \frac{263.6250}{27} = \boxed{9.7639}$$

F-Computed

$$\text{Block F-computed (R Fc)} = \frac{RMS}{E(a)MS} = \frac{43.2222}{27.8681} = \boxed{1.55}$$

$$\text{Factor A F-computed (A Fc)} = \frac{AMS}{E(a)MS} = \frac{84.1458}{27.8681} = \boxed{3.02}$$

$$\text{Factor B F-computed (B Fc)} = \frac{BMS}{E(b)MS} = \frac{123.8333}{9.7639} = \boxed{12.88}$$

$$\text{A x B F-computed (A x B Fc)} = \frac{ABMS}{E(b)MS} = \frac{30.5625}{9.7639} = \boxed{3.13}$$

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	3	129.6667	43.2222	1.55 ns	9.78	4.76
Main Plot (A)	2	166.2917	84.1458	3.02 ns	10.92	5.14
Error (a)	6	167.2083	27.8681			
Subplot (B)	3	371.5000	123.8333	12.68 **	4.50	2.96
A x B	6	183.3750	30.5625	3.13 *	3.56	2.46
Error (b)	27	263.6250	9.7639			
TOTAL	47	1,283.6667				

Coefficient of Variation (C.V.) may be computed in two ways: pooled C.V. or individual C.V. (one for A and one for B). Pooled (or combined) C.V. is computed by computing the average MSE. Average MSE may not be computed using arithmetic mean because the divisors (Error df) have different values. Weighted mean is therefore used:

$$\text{Pooled MSE} = \frac{167.2083 + 263.6250}{6 + 27} = \boxed{13.0556}$$

$$\text{Pooled C.V.} = \frac{\sqrt{\text{pooled MSE}}}{\text{Mean}} \times 100 = \frac{\sqrt{13.0556}}{22.4} \times 100 = \boxed{16.1\%}$$

Two separate C.V. (one for Factor A and another for Factor B & interaction) may also be computed:

$$\text{C.V. (a)} = \frac{\sqrt{\text{MSEa}}}{\text{Mean}} \times 100 = \frac{\sqrt{27.8681}}{22.4} \times 100 = \boxed{31.4\%}$$

$$\text{C.V. (b)} = \frac{\sqrt{\text{MSEb}}}{\text{Mean}} \times 100 = \frac{\sqrt{9.7639}}{22.4} \times 100 = \boxed{13.9\%}$$

The results of analysis of split-plot in CRD as well as in RCBD showed that no significant difference were noted among Factor A means compared to the highly significant differences noted when using factorial (in either CRD or RCBD). This is due to the partitioning of the error term into two components due to the main plot A and due to subplot B and the interaction between A and B. This resulted to high E(a)MS causing the main plot A means to be not significant.

Highly significant differences among means of Factor B similar to the results obtained in factorial however, note the increase in Fc value in split-plot compared to factorial.

Significant differences were noted among the interaction means between A and B as compared to no significant interaction when using factorial. This is because the partitioning of the error term resulted to a very small error (b) term which can detect slight variation in the means of B or AB.

The coefficient of variation (C.V.) has two values: one due to the main plot and another due to the subplot. Note that the C.V. of the subplot is much lower than that of the main plot, an indication that the subplot has a higher degree of precision in terms of detecting differences among treatment means. Note also that there was a decrease in C.V.(a) in split-plot in RCBD compared to that in split-plot in CRD. This is because in split-plot in CRD, the variation due to the block is confounded or added to the main effect (Error A).

Mean Comparison using LSD

A x B Mean Summary Table

	B₁	B₂	B₃	B₄	A Means
A₁	25.5	25.0	28.0	21.9	25.1
A₂	26.3	19.5	20.3	18.0	21.0
A₃	28.8	17.3	19.8	19.0	21.2
B Means	26.8	20.6	22.7	19.6	

Since Factor A levels were detected to be not significant, there is no need to compute for its LSD. On the other hand, Factor B levels were detected to be highly significant so either LSD 1% or LSD 5% may be used for pair comparison. When computed, the values are:

$$\text{LSD}_{0.05, 27 \text{ df}} = 2.031 \times 1.275 = \boxed{2.62}$$

$$\text{LSD}_{0.01, 27 \text{ df}} = 2.727 \times 1.275 = \boxed{3.53}$$

Any of the two LSD values will be used to compare the Treatment Means of the 3 Factor B levels depending on the desired probability level. Since there are 4 treatments, only 3 valid pair comparisons may be made. Comparing B₁ with B₂, B₃ and B₄ shows that at 1% probability level, B₁ is highly significantly different from all 3 treatment means.

	Mean	Abs. diff.
B ₁ =	26.8	
B ₂ =	20.6	6.2 **
B ₃ =	22.7	4.1 **
B ₄ =	19.6	7.2 **

Mean Comparison using DMRT

1. Comparison among main-plot treatment means. Since main-plot levels were detected to be not significant, there is no need to compute for its DMRT.
2. Comparison among subplot treatment means averaged over all main-plot treatments. The same procedure as in previous example in Factorial CRD is applicable using the appropriate s_d value. When computed, the Rp values for comparison of subplot treatment means will be:

$$s_d = \mathbf{9.7639}$$

$$\text{Error df} = \mathbf{27}$$

P	rp	Rp
2	2.915	4.41
3	3.050	4.57
4	3.135	4.64

And the mean comparison will be:

B ₁	= 26.8 a
B ₄	= 22.4 b
B ₂	= 20.6 b
B ₃	= 19.6 b

3. Comparison among subplot treatment means at the same main-plot treatment. Similar to the procedure in LSD, comparison is done among the 4 B levels at A₁, 4 B levels at A₂, and 4 B levels at A₃.

To illustrate the comparison of B₁, B₂, B₃ & B₄ at A₂ using the computed s_d value of 2.20, the computed Rp values are:

$$s_d = \mathbf{9.7639}$$

$$\text{Error df} = \mathbf{27}$$

P	rp	Rp
2	2.915	4.41
3	3.050	4.57
4	3.135	4.64

And the mean comparison will be

$$B_1 = 26.3 \text{ a}$$

$$B_3 = 20.3 \text{ b}$$

$$B_2 = 19.5 \text{ b}$$

$$B_4 = 18.0 \text{ b}$$

The same procedure may be applied to compare B_1 - B_4 at A_1 and at A_3 .

4. Comparison among main-plot means at the same subplot treatments. Comparison may be done separately among 3 levels of A at B_1 , at B_2 , at B_3 and at B_4 . Using the s_d value derived from formula in Item 4 for LSD computation (2.81), the R_p values can be computed.

$$2 = 8.16$$

$$3 = 8.57$$

To compare A_1 , A_2 & A_3 at B_3 :

$$A_1 = 28.0 \text{ a}$$

$$A_2 = 20.3 \text{ b}$$

$$A_3 = 19.8 \text{ b}$$

STRIP-PLOT (SPLIT-BLOCK) IN RANDOMIZED COMPLETE BLOCK DESIGN

Application

Strip-plot in RCBD is most appropriate in experiments dealing with two factors which are more or less equally important while the interaction is more important. This is also applicable if the two factors can be applied easily using large plots than using small plots. A typical example is a combination of irrigation system and frequency of spraying of pesticide. In this design, one treatment is applied in horizontal position while the other treatment in vertical position. As such, both Factor A and Factor B cover large plots while the interaction is much smaller than any of the two factors. The degree of precision is more or less equal for both Factors A and B but that of the interaction is more precise than any of the two factors.

Randomization

Randomization of Factor A is done independent of the randomization of Factor B but both will be randomized within each block. The procedure is as follows:

1. Determine the number of treatments in Factor A.
2. For Block 1, randomize the treatments in factor A in one direction (vertical or horizontal) using a randomization scheme described in RCBD. Note that all treatments must appear together as strips in one block, that is, use of two sub-blocks is not possible.

BLOCK 1

A_3	A_2	A_1

- Determine the number of treatments in Factor B.
- Randomize the number of treatments in Factor B by the same randomization scheme used in Factor A but in the other direction.

BLOCK 1

B_1		
B_3		
B_4		
A_3B_4	A_2	A_1

When each cell is completed, the combination of A and B will be:

BLOCK 1

A_3B_1	A_2B_1	A_1B_1
A_3B_3	A_2B_3	A_1B_3
A_3B_2	A_2B_2	A_1B_2
A_3B_4	A_2B_4	A_1B_4

- Follow Steps 1 to 4 for the succeeding blocks.

Following the randomization scheme above, a possible lay-out for strip-plot in RCBD with three levels of Factor A and four levels of Factor B with four replications is shown if the total area is very long and the fertility gradient is irregular or not known.

BLOCK 1

A_1B_1	A_3B_1	A_2B_1
A_1B_3	A_3B_3	A_2B_3
A_1B_2	A_3B_2	A_2B_2
A_1B_4	A_3B_4	A_2B_4

BLOCK 2

A_2B_3	A_1B_3	A_3B_3
A_2B_2	A_1B_2	A_3B_2
A_2B_4	A_1B_4	A_3B_4
A_2B_1	A_1B_1	A_3B_1

BLOCK 3

A_3B_4	A_2B_4	A_1B_4
A_3B_1	A_2B_1	A_1B_1
A_3B_2	A_2B_2	A_1B_2
A_3B_3	A_2B_3	A_1B_3

BLOCK 4

A_3B_2	A_2B_2	A_1B_2
A_3B_1	A_2B_1	A_1B_1
A_3B_4	A_2B_4	A_1B_4
A_3B_3	A_2B_3	A_1B_3

:

The lay-out may also be like this:

BLOCK 3	A ₁ B ₁	A ₃ B ₁	A ₂ B ₁	A ₂ B ₃	A ₁ B ₃	A ₃ B ₃	BLOCK 4
	A ₁ B ₂	A ₃ B ₂	A ₂ B ₂	A ₂ B ₂	A ₁ B ₂	A ₃ B ₂	
	A ₁ B ₃	A ₃ B ₃	A ₂ B ₃	A ₂ B ₄	A ₁ B ₄	A ₃ B ₄	
	A ₁ B ₄	A ₃ B ₄	A ₂ B ₄	A ₂ B ₁	A ₁ B ₁	A ₃ B ₁	
BLOCK 1	A ₃ B ₄	A ₂ B ₄	A ₁ B ₄	A ₃ B ₂	A ₂ B ₂	A ₁ B ₂	BLOCK 2
	A ₃ B ₁	A ₂ B ₁	A ₁ B ₁	A ₃ B ₁	A ₂ B ₁	A ₁ B ₁	
	A ₃ B ₂	A ₂ B ₂	A ₁ B ₂	A ₃ B ₄	A ₂ B ₄	A ₁ B ₄	
	A ₃ B ₃	A ₂ B ₃	A ₁ B ₃	A ₃ B ₃	A ₂ B ₂	A ₁ B ₃	

It is not advisable to use very long plots because it will result to one of the factors having very long area coverage which might result to high variability within the block. As such, square or nearly square plots are recommended.

The same set of data used in previous examples will be used to have a comparison between Split-Plot and Strip-Plot in RCBD.

FACTOR A	FACTOR B	REP I	REP II	REP III	REP IV	TOTAL	MEAN
A1	B ₁	28	29	23	22	102	25.5
A1	B ₂	25	24	27	24	100	25.0
A1	B ₃	26	28	31	27	112	28.0
A1	B ₄	23	24	17	23	87	21.8
A2	B ₁	27	31	20	27	105	26.2
A2	B ₂	16	20	22	20	78	19.5
A2	B ₃	17	23	16	25	81	20.2
A2	B ₄	19	18	17	18	72	18.0
A3	B ₁	30	29	23	33	115	28.8
A3	B ₂	16	14	16	23	69	17.2
A3	B ₃	18	16	20	25	79	19.8
A3	B ₄	17	20	10	29	76	19.0
TOTAL		262	276	242	296	1,076	22.4

Strip-plot will require all the three 2-way tables: AxB, AxR, and BxR

A x B SUMMARY TABLE

	B ₁	B ₂	B ₃	B ₄	A TOTALS	A MEANS
A ₁	102	100	112	87	401	25.1
A ₂	105	78	81	72	336	21.0
A ₃	115	69	79	76	339	21.2
B TOTALS	322	247	272	235	1,076	
B MEANS	26.8	20.6	22.7	19.6		

A x R SUMMARY TABLE

	R ₁	R ₂	R ₃	R ₄	A TOTALS	A MEANS
A ₁	102	105	98	96	401	25.1
A ₂	79	92	75	90	336	21.0
A ₃	81	79	69	110	339	21.2
R TOTALS	262	276	242	296	1,076	

B x R SUMMARY TABLE

	R ₁	R ₂	R ₃	R ₄	B TOTALS	B MEANS
B ₁	85	89	66	82	322	26.8
B ₂	57	58	65	67	247	20.6
B ₃	61	67	67	77	272	22.7
B ₄	59	62	44	70	235	19.6
R TOTALS	262	276	242	296	1,076	

Basic values

a = number or levels of Factor A	= 3
b = number or levels of Factor B	= 4
r = number of replications	= 4
GT = (A ₁ B ₁ R ₁) + (A ₁ B ₁ R ₂) + . . . + (A ₃ B ₄ R ₄) = 26 + 29 + . . . + 29	= 1076
$\bar{X} = \frac{(GT)}{abr} = \frac{1,076}{(3 \times 4 \times 4)}$	= 22.4

Degrees of Freedom

Block	= (r - 1)	= (4 - 1)	= 3
Factor A	= (a - 1)	= (3 - 1)	= 2
Error (a)	= (a - 1)(r - 1)	= (3 - 1)(4 - 1)	= 6
Factor B	= (b - 1)	= (4 - 1)	= 3
Error (b)	= (b - 1)(r - 1)	= (4 - 1)(4 - 1)	= 9
A x B	= (a - 1)(b - 1)	= (3 - 1)(4 - 1)	= 6
Error (ab)	= (a - 1)(b - 1)(r - 1)	= (3 - 1)(4 - 1)(4 - 1)	= 18
Total	= (abr - 1)	= 3 x 4 x 4 - 1	= 47

Error (a) will be used to test the significance of Factor A and Block, Error (b) will be used to test the significance of Factor B, and Error (ab) will test the significance of A x B.

Squares

$$C.F. = \frac{GT^2}{abr} = \frac{1076^2}{3 \times 4 \times 4} = \boxed{24,120.3333}$$

$$\frac{\Sigma R^2}{ab} = \frac{R_1^2 + R_2^2 + \dots + R_4^2}{ab} = \frac{262^2 + 276^2 + \dots + 296^2}{3 \times 4} = \boxed{24,250.0000}$$

$$\frac{\Sigma A^2}{br} = \frac{A_1^2 + A_2^2 + A_3^2}{br} = \frac{401^2 + 336^2 + 330^2}{4 \times 4} = \boxed{24,266.6250}$$

$$\frac{\Sigma B^2}{ar} = \frac{B_1^2 + B_2^2 + B_3^2 + B_4^2}{ar} = \frac{322^2 + 247^2 + 272^2 + 235^2}{3 \times 4} = \boxed{24,491.8333}$$

$$\frac{\Sigma \Sigma (AB)^2}{r} = \frac{(A_1B_1)^2 + (A_1B_2)^2 + \dots + (A_3B_4)^2}{r} = \frac{102^2 + 100^2 + \dots + 76^2}{3} = \boxed{24,843.5000}$$

$$\frac{\Sigma \Sigma (AR)^2}{b} = \frac{(A_1R_1)^2 + (A_1R_2)^2 + \dots + (A_3R_4)^2}{b} = \frac{102^2 + 105^2 + \dots + 110^2}{4} = \boxed{24,585.5000}$$

$$\frac{\Sigma \Sigma (BR)^2}{a} = \frac{(B_1R_1)^2 + (B_1R_2)^2 + \dots + (B_3R_4)^2}{a} = \frac{102^2 + 100^2 + \dots + 75^2}{3} = \boxed{24,288.6250}$$

$$\Sigma \Sigma \Sigma (ABR)^2 = (A_1B_1R_1)^2 + (A_1B_1R_2)^2 + \dots + (A_3B_4R_4)^2 = 26^2 + 29^2 + \dots + 29^2 = \boxed{25,404.0000}$$

Sums of Squares

$$ToSS = \Sigma \Sigma (ABR)^2 - C.F. = 25,404.0000 - 24,120.3333 = \boxed{1,283.6667}$$

$$RSS = \frac{\Sigma R^2}{ab} - C.F. = 24,250.0000 - 24,120.3333 = \boxed{129.6667}$$

$$ASS = \frac{\Sigma A^2}{br} - C.F. = 24,266.6250 - 24,120.3333 = \boxed{168.2917}$$

$$E(a) SS = \frac{\Sigma \Sigma (AR)^2}{b} - \frac{\Sigma A^2}{br} = 24,585.5000 - 24,288.6250 = \boxed{296.8750}$$

$$E(a) SS = \frac{\Sigma \Sigma (AR)^2}{b} - \frac{\Sigma A^2}{br} - \frac{\Sigma R^2}{ab} + C.F. = 24,585.5000 - 24,288.6250 - 24,250.0000 + 24,120.3333 = \boxed{167.2083}$$

$$BSS = \frac{\Sigma B^2}{ar} - C.F. = 24,491.8333 - 24,120.3333 = \boxed{371.5000}$$

$$E(b)SS = \frac{\Sigma\Sigma(BR)^2}{a} - \frac{\Sigma B^2}{ar} - \frac{\Sigma R^2}{ab} + C.F.$$

$$= 24,780.6667 - 24,491.98333 - 24,250.000 + 24,120.3333 = \boxed{159.1667}$$

$$ABSS = \frac{\Sigma\Sigma(AB)^2}{r} - \frac{\Sigma A^2}{br} - \frac{\Sigma B^2}{ar} + C.F.$$

$$= 24,843.5000 - 24,288.6250 - 24,491.98333 + 24,120.3333 = \boxed{183.3750}$$

$$E(ab)SS = \Sigma\Sigma\Sigma(ABR)^2 - \frac{\Sigma\Sigma(AB)^2}{r} - \frac{\Sigma\Sigma(AR)^2}{b} - \frac{\Sigma\Sigma(BR)^2}{a} + \frac{\Sigma(A)^2}{br} + \frac{\Sigma(B)^2}{ar} + \frac{\Sigma(R)^2}{ab} - C.F.$$

$$= 25404.0000 - 24843.5000 - 24585.5000 - 24288.6250 + 24288.6250 + 24491.8333 + 24250.000 - 24120.3333$$

$$= \boxed{104.4583}$$

Mean Squares

$$\text{Block Mean Square (RMS)} = \frac{RSS}{R \text{ df}} = \frac{129.6667}{3} = \boxed{43.2222}$$

$$\text{Factor A Mean Square (AMS)} = \frac{ASS}{A \text{ df}} = \frac{168.2917}{2} = \boxed{84.1458}$$

$$\text{Error (a) Mean Square (EaMS)} = \frac{E(a)SS}{E(a) \text{ df}} = \frac{167.2083}{6} = \boxed{27.8681}$$

$$\text{Factor B Mean Square (BMS)} = \frac{BSS}{B \text{ df}} = \frac{371.5000}{3} = \boxed{123.8333}$$

$$\text{Error (b) Mean Square (EbMS)} = \frac{E(b)SS}{E(b) \text{ df}} = \frac{159.1667}{9} = \boxed{17.6852}$$

$$A \times B \text{ Mean Square (ABMS)} = \frac{ABSS}{AB \text{ df}} = \frac{183.3750}{6} = \boxed{30.5625}$$

$$\text{Error (ab) Mean Square (EabMS)} = \frac{E(ab)SS}{E(ab) \text{ df}} = \frac{104.4583}{18} = \boxed{5.8032}$$

F-Computed

$$\text{Factor A F-computed (A Fc)} = \frac{AMS}{E(a)MS} = \frac{84.1458}{27.8681} = \boxed{3.02}$$

$$\text{Factor B F-computed (B Fc)} = \frac{BMS}{E(b)MS} = \frac{123.8333}{17.6852} = \boxed{7.00}$$

$$A \times B \text{ F-computed (A x B Fc)} = \frac{ABMS}{E(ab)MS} = \frac{30.5625}{5.8032} = \boxed{5.27}$$

ANALYSIS OF VARIANCE

Source	df	SS	MS	Fc	Ft 1%	Ft 5%
Block	3	129.6667	43.2222			
Factor A	2	166.2917	84.1458	3.02 ns	10.92	5.14
Error (a)	6	167.2083	27.8681			
Factor B	3	371.5000	123.8333	7.00 **	6.99	3.86
Error (b)	9	159.1667	17.6852			
A x B	6	183.3750	30.5625	5.27 **	4.01	2.46
Error (ab)	18	104.4583	5.8032			
TOTAL	47	1,283.6667				

Coefficient of Variation (C.V.) may be computed in two ways: pooled C.V. or individual C.V. (one for A and one for B). Pooled (or combined) C.V. is computed by computing the average MSE. Average MSE may not be computed using arithmetic mean because the divisors (Error df) have different values. Weighted mean is therefore used:

$$\text{Pooled MSE} = \frac{167.2083 + 159.1667 + 104.4583}{6 + 9 + 18} = \boxed{13.0556}$$

$$\text{Pooled C.V.} = \frac{\sqrt{\text{pooled MSE}}}{\text{Mean}} \times 100 = \frac{\sqrt{13.0556}}{22.4} \times 100 = \boxed{16.1\%}$$

Three individual C.V. may be computed as:

$$\text{C.V. (a)} = \frac{\sqrt{\text{MSE(a)}}}{\text{Mean}} \times 100 = \frac{\sqrt{27.8681}}{22.4} \times 100 = \boxed{31.4\%}$$

$$\text{C.V. (b)} = \frac{\sqrt{\text{MSE(b)}}}{\text{Mean}} \times 100 = \frac{\sqrt{17.6852}}{22.4} \times 100 = \boxed{18.8\%}$$

$$\text{C.V. (ab)} = \frac{\sqrt{\text{MSE(ab)}}}{\text{Mean}} \times 100 = \frac{\sqrt{5.8032}}{22.4} \times 100 = \boxed{10.7\%}$$

When the same data used for factorial and split-plot were analyzed using strip-plot, the results showed that mean differences among Factor A are not significant similar to the result in split-plot.

Mean differences among Factor B levels are both highly significant in split-plot and strip plot but it is noticeable that the Fc for split-plot is much higher than that in strip-plot. This is expected because in strip-plot, Factor B is as important as Factor A while in split-plot, Factor B is more important than Factor A.

The big difference lies in the interaction between A and B where it is highly significant in strip-plot and significant only in split-plot. This is due to the further partitioning of the Error (b) from split-plot into two components in strip-plot: one due to main effect B and another due to the interaction between A and B. This resulted to a much lower Error Mean Square (9.7639 in split-plot compared to 5.8032 in strip-plot) which is more precise in detecting significant differences among the means of the interaction.

Mean Comparison using LSD

A x B Mean Summary Table

	B ₁	B ₂	B ₃	B ₄	A Means
A ₁	25.5	25.0	28.0	21.9	25.1
A ₂	26.3	19.5	20.3	18.0	21.0
A ₃	28.8	17.3	19.8	19.0	21.2
B Means	26.8	20.6	22.7	19.6	

There are 3 possible kinds of pair comparisons in Strip-plot:

1. Between two Factor A treatment means averaged over all Factor B treatments. Since Factor A levels were detected to be not significant, there is no need to compute for its LSD. In case they are significantly different, the formula for standard error of mean difference (s_d) is:

$$s_d = \sqrt{\frac{2 \times \text{MSE}(a)}{b \times r}}$$

and LSD is computed by multiplying the s_d value with t value at 5% probability level and at Error(a) df.

2. Between two Factor B treatment means averaged over all Factor A treatments. The same procedure as in LSD for Factorial CRD is applicable. When computed, the values are:

$$s_d = \sqrt{\frac{2 \times \text{MSE}(b)}{a \times r}} = \sqrt{\frac{2 \times 17.6852}{3 \times 4}} = \boxed{1.72}$$

$$\text{LSD}_{0.05, 9 \text{ df}} = 2.262 \times 1.72 = \boxed{3.88}$$

$$\text{LSD}_{0.01, 9 \text{ df}} = 3.250 \times 1.72 = \boxed{5.58}$$

Any of the two LSD values may be used to compare the Treatment Means of the 3 Factor B levels depending on the desired probability level. Since there are 4 treatments, only 3 valid pair comparisons may be made. Comparing B₁ with B₂, B₃ and B₄ shows that at 1% probability level, B₁ is highly significantly different from all 3 treatment means.

	Mean	Abs. diff.
B ₁ =	26.8	
B ₂ =	20.6	6.2 **
B ₃ =	22.7	4.1 *
B ₄ =	19.6	7.2 **

3. Comparison between two Factor A treatment means at the same Factor B treatment or between two Factor B treatment means at the same Factor A treatment. When the interaction between two factors is detected to be significant or highly significant, differences among Factor A and Factor B treatment means are ignored and the Interaction A x B treatment means are compared. In the example, 7 separate group comparisons will be made :

- 4 B levels at A₁
- 4 B levels at A₂
- 4 B levels at A₃
- 3 A levels at B₁
- 3 A levels at B₂
- 3 A levels at B₃
- 3 A levels at B₄

The formula is:

$$s_d = \sqrt{\frac{2 \times E(ab)}{r}} = \sqrt{\frac{2 \times 5.8032}{4}} = 1.59$$

$$LSD_{0.05, 18 \text{ df}} = 2.101 \times 1.59 = 3.35$$

$$LSD_{0.01, 18 \text{ df}} = 2.878 \times 1.59 = 4.49$$

With 4 treatments, only 3 valid comparisons can be done. If B₁ will be compared to B₂, B₃, & B₄,

	Mean	Abs. diff.
B ₁ =	25.5	
B ₂ =	25.0	0.5 ns
B ₃ =	28.0	3.0 ns
B ₄ =	21.9	3.6 *

B₁ is not significantly different from B₂ and B₃ but significantly different from B₄.

Mean Comparison using DMRT

1. Among main-plot treatment means. Since main-plot levels were detected to be not significant, there is no need to compute for its DMRT.
2. Among Factor B treatment means averaged over all Factor A treatments. The same procedure as in previous example in Factorial CRD is applicable using the appropriate s_d value. When computed, the Rp values for comparison of subplot treatment means will be:

$$\begin{aligned} R_p \\ 2 &= 3.89 \\ 3 &= 4.06 \\ 4 &= 4.14 \end{aligned}$$

And the mean comparison will be:

$$\begin{aligned} B_1 &= 26.8 \text{ a} \\ B_4 &= 22.4 \text{ b} \\ B_2 &= 20.6 \text{ b} \\ B_3 &= 19.6 \text{ b} \end{aligned}$$

- 3a. Comparison among Factor A treatment means at the same Factor B treatment. Comparison may be done separately among 3 levels of A at B₁, at B₂, at B₃ and at B₄. Using the s_d value derived from formula in Item 4 for LSD computation (2.84), the Rp values can be computed.

$$\begin{aligned} 2 &= 5.83 \\ 3 &= 6.12 \end{aligned}$$

To compare A₁, A₂ & A₃ at B₁:

$$\begin{aligned} A_1 &= 25.5 \text{ a} \\ A_2 &= 26.3 \text{ a} \\ A_3 &= 28.8 \text{ a} \end{aligned}$$

To compare A₁, A₂ & A₃ at B₃:

$$\begin{aligned} A_1 &= 28.0 \text{ a} \\ A_2 &= 20.3 \text{ b} \\ A_3 &= 19.8 \text{ b} \end{aligned}$$

- 3b. Comparison among Factor B treatment means at the same Factor A treatment. Similar to the procedure in LSD, comparison is done among the 4 B levels at A₁, 4 B levels at A₂, and 4 B levels at A₃.

A x B Mean Summary Table

	B₁	B₂	B₃	B₄	A Means
A₁	25.5	25.0	28.0	21.9	25.1
A₂	26.3	19.5	20.3	18.0	21.0
A₃	28.8	17.3	19.8	19.0	21.2
B Means	26.8	20.6	22.7	19.6	

To illustrate the comparison of B₁, B₂, B₃ & B₄ at A₁ using the computed s_d value of 2.79, the computed Rp values are:

$$2 = 5.73$$

$$3 = 6.02$$

$$4 = 6.18$$

And the mean comparison will be

$$B_1 = 25.5 \text{ ab}$$

$$B_2 = 25.0 \text{ ab}$$

$$B_3 = 28.0 \text{ a}$$

$$B_4 = 21.9 \text{ b}$$

The same procedure may be applied to compare B₁-B₄ at A₂ and at A₃.

COMPARISON OF THE THREE EXTENSIONS OF THE BASIC DESIGN

It is worth looking at the basic differences in the analysis of variance of the extensions of the basic design. The degrees of freedom and sums of squares show that the replication, Factor A and Factor B did not change in value and that only the error term changes from one design to another: In factorial, there is only one error term to test A, B and A x B. In split-plot, the error term is divided into two components: error (a) to test A and error (b) to test B and A x B. Comparing split-plot and strip-plot, error (a) value did not change while error (b) in split-plot was divided into two components in strip-plot: error (b) to test B and error (c) to test A x B.

In Completely Randomized Design

Source	df		SS	
	Fact.	SpP	Fact.	SpP
Factor A	2	2	168.2917	168.2917
Error	-	9	-	296.8750
Factor B	3	3	371.5000	371.5000
A x B	6	6	183.3750	183.3750
Error	36	27	560.5000	263.6250
TOTAL	47	47	1,283.6667	1,283.6667

In Randomized Complete Block Design:

Source	df			SS		
	Fact.	SpP	StP	Fact.	SpP	StP
Block	3	3	3	129.6667	129.6667	129.6667
Factor A	2	2	2	166.2917	166.2917	166.2917
Error	-	6	6	-	167.8750	167.8750
Factor B	3	3	3	371.5000	371.5000	371.5000
Error						159.1667
A x B	6	6	6	183.3750	183.3750	183.3750
Error	36	27	18	560.5000	263.6250	104.4583
TOTAL	47	47	47	1,283.6667	1,283.6667	1,283.6667

The following observations may be noted when comparing the sum of squares of the different designs:

- In Factorial, the error is pooled for Factor A, Factor B, and the interaction between A and B factors. This is so because the factor combinations (A x B) are randomized within each replication so the size of A, B and AB are the same.
- In Split-plot, the error term is partitioned into two components: Error (a) which is used to test Factor A, and Error (b) which is used to test Factor B and the interaction between A and B. This is so because Factor A is randomized within each replicate while Factor B is randomized within each Factor A within each replicate. In terms of plot size, A is bigger than B
- In Strip-plot, the error term is partitioned into three components: Error (a) which is used to test Factor A, Error (b) which is used to test Factor B, and Error (c) which is used to test the interaction between A and B. This is so because both factors A and B are randomized independently within each replication.

Interpretation of results for Factorial, Split-plot and Strip-plot

1. In an experiment involving two or more factors, it is always important to look first at the interaction (A x B).
 - If the Fc value of the interaction is significant and greater than the Fc value of either main factor, ignore the significance of the main factors and concentrate on the interaction.
 - Even if the differences among means of A factor are significant, if the Fc value of the interaction is much greater, any variation detected among means of in Factor A cannot be attributed entirely to the effect of factor A alone as the variation may have been caused partly by Factor B.
 - Conversely, even if the differences among means of B factor are significant, if the Fc value of the interaction is much greater, any variation detected among means in Factor B cannot be attributed entirely to the effect of Factor B alone as the variation may have been caused partly by Factor A.
2. If the Fc value of Factor A is much greater than that of the interaction, ignore the significance of the interaction and consider the variation in Factor A as due to the effect of the main treatment. The same rule applies for Factor B.

A summary of the possible interpretation of the results of two-factor experiment is shown in the next table.

A summary table of the interpretation of possible results of a two-factor designs may be useful:

Factor A	Factor B	A x B	INTERPRETATION
ns	ns	ns	- Since <u>Factor A</u>, <u>Factor B</u> and <u>Interaction AxB</u> are all not significant, the conclusion/recommendation will depend on the type of experiment being conducted and the levels or kinds of treatments in each factor. - If the factor consists of different varieties, it indicates the new varieties are not significantly different from the check variety so it is still better to use the check variety. However, other traits/characters gathered should be considered in making the final recommendation. - If the factor consists of different levels of fertilizer, pesticide or chemical, or different seed rates, the lowest level or rate may be recommended since it will mean less cost or input for the farmers. - No specific <u>AxB</u> combination can be concluded significantly different from the others, however, a specific recommendation may still be possible if the two factors involve varieties and levels of fertilizer, pesticide or chemical, or different seed rates.
* or **	ns	ns	- When <u>A</u> is significant or highly significant while <u>B</u> and <u>AxB</u> are not significant, the best level or kind of <u>A</u> is selected. - If <u>B</u> consists of different levels of fertilizer, pesticide or chemical, or different seed rates, the lowest level or rate may be recommended since it will mean less cost or input for farmers. - A specific <u>AxB</u> combination may also be selected if <u>B</u> consists of different levels of fertilizer, pesticide or chemical, or different seed rates so the lowest level or rate combined with the best <u>A</u> level or kind may be recommended since it will mean less cost or input for farmers.
ns	* or **	ns	- When <u>B</u> is significant or highly significant while <u>A</u> and <u>AxB</u> are not significant, the best level or kind of <u>B</u> is selected. - If <u>A</u> consists of different levels of fertilizer, pesticide or chemical, or different seed rates, the lowest level or rate may be recommended since it will mean less cost of input for farmers. - A specific <u>AxB</u> combination may also be selected if <u>A</u> consists of different levels of fertilizer, pesticide or chemical, or different seed rates so the lowest level or rate combined with the best <u>B</u> level or kind may be recommended since it will mean less cost or input for farmers.
* or **	* or **	ns	- When both <u>A</u> and <u>B</u> are significant (or highly significant) while <u>AxB</u> is not significant, the best level or kind of <u>A</u> and <u>B</u> are selected. - If either <u>A</u> or <u>B</u> consists of different levels of fertilizer, pesticide or chemical, or different seed rates, a specific <u>AxB</u> combination may still be selected if it will mean lower cost or input.
ns	ns	* or **	Since both <u>A</u> and <u>B</u> are not significant while <u>AxB</u> is significant (or highly significant), the best <u>AxB</u> combination is recommended.
*	ns	* or **	Since <u>AxB</u> is significant (or highly significant), the best <u>AxB</u> combination is recommended while the result of <u>A</u> is ignored, that is, no specific <u>A</u> level or kind may be recommended.
ns	*	* or **	Since <u>AxB</u> is significant (or highly significant), the best <u>AxB</u> combination is recommended while the result of <u>B</u> is ignored even if it is significant, that is, no specific <u>B</u> level or kind may be recommended.

**	ns	*	- If <u>A</u> is highly significant and <u>AxB</u> is just significant, <u>AxB</u> may be ignored so a specific level or kind of <u>A</u> may be recommended. - No specific level or kind of <u>B</u> should be recommended. - It is still possible to recommend a specific <u>AxB</u> combination.
ns	**	*	- If <u>B</u> is highly significant and <u>AxB</u> is just significant, <u>AxB</u> may be ignored so a specific level or kind of <u>B</u> may be recommended. - No specific level of <u>A</u> should be recommended. - It is still possible to recommend a specific <u>AxB</u> combination.
*	*	*	- If <u>A</u>, <u>B</u>, and <u>AxB</u> are all significant, <u>IGNORE</u> both <u>A</u> and <u>B</u>, and <u>CONCENTRATE</u> only on the <u>AxB</u>. - A specific <u>AxB</u> combination may be recommended.
**	*	*	- If <u>A</u> is highly significant while <u>B</u> and <u>AxB</u> are significant, <u>AxB</u> may be ignored (in relation to <u>A</u>) so a specific level or kind of <u>A</u> may be recommended. - No specific level of <u>B</u> should be recommended. - A specific <u>AxB</u> combination may be recommended.
*	**	*	- If <u>B</u> is highly significant while <u>A</u> and <u>AxB</u> are significant, <u>AxB</u> may be ignored (in relation to <u>B</u>) so a specific level or kind of <u>B</u> may be recommended. - No specific level of <u>A</u> should be recommended. - A specific <u>AxB</u> combination may be recommended.
**	**	*	- If both <u>A</u> and <u>B</u> are highly significant while <u>AxB</u> is significant, <u>AxB</u> may be ignored (in relation to either <u>A</u> or <u>B</u>) so a specific level or kind <u>A</u> and <u>B</u> are recommended. - A specific <u>AxB</u> combination may be recommended.
**	*	**	- If both <u>A</u> and <u>AxB</u> are highly significant while <u>B</u> is significant, <u>IGNORE</u> both <u>A</u> and <u>B</u> and concentrate on <u>AxB</u> so a specific <u>AxB</u> combination may be recommended. - If the Fc value of <u>A</u> is much, much higher (very highly significant), a specific level of <u>A</u> may be recommended.
*	**	**	- If both <u>B</u> and <u>AxB</u> are highly significant while <u>A</u> is significant, <u>IGNORE</u> both <u>A</u> and <u>B</u> and concentrate on <u>AxB</u> so a specific <u>AxB</u> combination may be recommended. - If the Fc value of <u>B</u> is much, much higher (very highly significant), a specific level of <u>B</u> may be recommended.
**	**	**	- This is similar to the case where <u>A</u>, <u>B</u>, and <u>AxB</u> are all significant so <u>IGNORE</u> both <u>A</u> and <u>B</u> and <u>CONCENTRATE</u> on <u>AxB</u> so a specific <u>AxB</u> combination may be recommended. - If <u>A</u>, <u>B</u> and <u>AxB</u> are all highly significant but the Fc value of either <u>A</u> or <u>B</u> is much, much higher (also called very highly significant) than the Fc value of <u>AxB</u>, a specific level or kind of <u>A</u> or specific level or kind of <u>B</u> may be selected together with the specific <u>AxB</u> combination.

Interpretation of results of three-factor experiments (triple Factorial, Factorial-split, Factorial-strip, Split-split, Split-factorial, Split-strip, Strip-factorial, Strip-split) is very complicated so as a general rule, IT IS NOT ADVISABLE to use these further extensions of basic designs UNLESS very necessary. There will be four (4) interactions (AxB, AxC, BxC, AxBxC) so recommendation will be very difficult especially if two or more of these interactions are detected to be significant.

FURTHER EXTENSIONS OF THE BASIC DESIGN

As a general rule, the use of complex designs is not advised unless very necessary, that is, the assessment of three or more factors simultaneously is absolutely needed. The difficulty in using the design lies in the following:

1. the experimental area will increase tremendously as the number of factors is increased so the cost of running the an experiment will have to be considered;
2. interpretation of results becomes more complicated since the number of interactions increases as the number of factors increases. For example, the interaction in two-factor experiment is only one, with three factors, the number of interactions is four, with four factors, the number of interactions is eight, and so on.

For theoretical purposes, the sums of squares of the extensions of basic designs can be computed by determining the degrees of freedom of the factors or factor combinations. The appropriate formula can be established by substituting the different terms of the expanded degrees of freedom with the appropriate sums of squares. In addition, the following principles should be considered:

1. **Factorial.** In factorial experiments, the factor combinations are always randomized in the experimental area in the case of CRD (or in each block in the case of RCBD) so the size of each factor and the interaction between or among them is always equal. Since they are all equal, only one error term is used to test for the significance of mean differences. The degrees of freedom for a 3-factor factorial in RCBD are:

Block	$(r-1) =$	$r - 1$
Factor A	$(a-1) =$	$a - 1$
Factor B	$(b-1) =$	$b - 1$
AxB	$(a-1)(b-1)=$	$ab - a - b + 1$
Factor C	$(c-1) =$	$c - 1$
AxC	$(a-1)(c-1)=$	$ac - a - c + 1$
BxC	$(b-1)(c-1)=$	$bc - b - c + 1$
AxBxC	$(a-1)(b-1)(c-1)=$	$abc - ab - ac - bc + a + b + c - 1$
Error	$(abc-1)(r-1) =$	$abcr - abc - r + 1$

For a four-factor factorial in RCBD, the corresponding degrees of freedom are as follows:

Block	$(r-1) =$	$r - 1$
Factor A	$(a-1) =$	$a - 1$
Factor B	$(b-1) =$	$b - 1$
A x B	$(a-1)(b-1) =$	$ab - a - b + 1$
Factor C	$(c-1) =$	$c - 1$
A x C	$(a-1)(c-1) =$	$ac - a - c + 1$
B x C	$(b-1)(c-1) =$	$bc - b - c + 1$
A x B x C	$(a-1)(b-1)(c-1) =$	$abc - ab - ac - bc + a + b + c - 1$
Factor D	$(d-1) =$	$d - 1$
A x D	$(a-1)(d-1) =$	$ad - a - d + 1$
B x D	$(b-1)(d-1) =$	$bd - b - d + 1$
C x D	$(c-1)(d-1) =$	$cd - c - d + 1$
A x B x D	$(a-1)(b-1)(d-1) =$	$abd - ab - ad - bd + a + b + d - 1$
A x C x D	$(a-1)(c-1)(d-1) =$	$acd - ac - ad - cd + a + c + d - 1$
B x C x D	$(b-1)(c-1)(d-1) =$	$bcd - bc - bd - cd + b + c + d - 1$
A x B x C x D	$(a-1)(b-1)(c-1)(d-1) =$	$abcd - abc - abd - acd - bcd + ab + ac + ad + bc + bd + cd - a - b - c - d + 1$
Error	$(abcd-1)(r-1) =$	$abcdr - abcd - r + 1$

2. Split-plot. The main feature of split-plot design is that the main plots are randomized in the experimental area in the case of CRD (or in each block in the case of RCBD), the sub-plots are randomized within each main plot, the sub-sub-plots are randomized within each sub-plot, and so on. Since the sub-plot is randomized within each main plot and not within the experimental area or block, the sub-plot (B) is considered nested within the main plot (A). Consider a split-split-plot design in RCBD. The appropriate degrees of freedom are as follows:

Block	$(r-1) =$	$r - 1$
Factor A	$(a-1) =$	$a - 1$
Error (a)	$(a-1)(r-1) =$	$ar - a - r + 1$
Factor B	$(b-1) =$	$b - 1$
AxB	$(a-1)(b-1) =$	$ab - a - b + 1$
Error (b)	$a(b-1)(r-1) =$	$abr - ab - ar + a$
Factor C	$(c-1) =$	$c - 1$
AxC	$(a-1)(c-1) =$	$ac - a - c + 1$
BxC	$(b-1)(c-1) =$	$bc - b - c + 1$
AXBxC	$(a-1)(b-1)(c-1) =$	$abc - ab - ac - bc + a + b + c - 1$
Error	$ab(c-1)(r-1) =$	$abcr - abc - abr + ab$

For a four-factor split (split-split-split-plot) in RCBD, the corresponding degrees of freedom are as follows:

Block	$(r-1) =$	$r - 1$
Factor A	$(a-1) =$	$a - 1$
Error (a)	$(a-1)(r-1) =$	$ar - a - r + 1$
Factor B	$(b-1) =$	$b - 1$
A x B	$(a-1)(b-1) =$	$ab - a - b + 1$
Error (b)	$a(b-1)(r-1) =$	$abr - ab - ar + a$
Factor C	$(c-1) =$	$c - 1$
A x C	$(a-1)(c-1) =$	$ac - a - c + 1$
B x C	$(b-1)(c-1) =$	$bc - b - c + 1$
A x B x C	$(a-1)(b-1)(c-1) =$	$abc - ab - ac - bc + a + b + c - 1$
Error (c)	$ab(c-1)(r-1) =$	$abcr - abc - abr + ab$
Factor D	$(d-1) =$	$d - 1$
A x D	$(a-1)(d-1) =$	$ad - a - d + 1$
B x D	$(b-1)(d-1) =$	$bd - b - d + 1$
C x D	$(c-1)(d-1) =$	$cd - c - d + 1$
A x B x D	$(a-1)(b-1)(d-1) =$	$abd - ab - ad - bd + a + b + d - 1$
A x C x D	$(a-1)(c-1)(d-1) =$	$acd - ac - ad - cd + a + c + d - 1$
B x C x D	$(b-1)(c-1)(d-1) =$	$bcd - bc - bd - cd + b + c + d - 1$
A x B x C x D	$(a-1)(b-1)(c-1)(d-1) =$	$abcd - abc - abd - acd - bcd + ab + ac + ad + bc + bd + cd - a - b - c - d + 1$
Error (d)	$abc(d-1)(r-1) =$	$abcdr - abcd - abcr + abc$

3. Strip-strip plot. Strip-plot can be extended up to 3 factors. Although it is a valid design, the physical lay-out is not possible (unless the field is 3-dimensional). The degrees of freedom for a strip-strip plot design in RCBD are as follows:

Block	$(r-1) =$	$r - 1$
Factor A	$(a-1) =$	$a - 1$
Error (a)	$(a-1)(r-1) =$	$ar - a - r + 1$
Factor B	$(b-1) =$	$b - 1$
Error (b)	$(b-1)(r-1) =$	$br - b - r + 1$
AxB	$(a-1)(b-1) =$	$ab - a - b + 1$
Error (ab)	$(a-1)(b-1)(r-1) =$	$abr - ab - ar - br + a + b + r - 1$
Factor C	$(c-1) =$	$c - 1$
Error (c)	$(c-1)(r-1) =$	$cr - c - r + 1$
AxC	$(a-1)(c-1) =$	$ac - a - c + 1$
Error (ac)	$(a-1)(c-1)(r-1) =$	$acr - ac - ar - cr + a + c + r - 1$
BxC	$(b-1)(c-1) =$	$bc - b - c + 1$
Error (bc)	$(b-1)(c-1)(r-1) =$	$bcr - bc - br - cr + b + c + r - 1$
AxBxC	$(a-1)(b-1)(c-1) =$	$abc - ab - ac - bc + a + b + c - 1$
Error (abc)	$(a-1)(b-1)(c-1)(r-1) =$	$abcr - abc - abr - acr - bcr + ab + ac + ar + bc + br + cr - a - b - c - r + 1$

Note that different combinations of any (or all) of factorial, split-plot, and strip-plot can also be made (although not recommended unless absolutely necessary). Some examples of extensions of basic designs in RCBD are:

- **Factorial-split** = A & B in factorial with C split into AB
- **Factorial-strip** = A & B in factorial, AB and C are in strip
- **Split-factorial** = Factors B & C in factorial and BC split into A
- **Split-strip** = Factors B & C are in strip; BC is split into A
- **Strip-split** = A & B are in strip plot with C split into AB
- **Split-strip-factorial** = A & BC are in factorial, B & C are in strip plot and D is split into BC

REGRESSION AND CORRELATION

When using simple ANOVA, only one variable is analyzed at any one time. With ANOCOVA, one variable is analyzed but the effect of another variable is removed. However, there are cases where one variable has an effect on another. In characterizing the association between characters, there is a need for statistical procedures that can simultaneously handle several variables. If two plant characters are measured to represent crop response, the analysis of variance and mean comparison procedures can evaluate only one character at a time, even though response in one character may affect the other, or treatment effects may simultaneously influence both characters. Regression and correlation analysis allows a researcher to examine any one or a combination of the three types of association described earlier provided that the variables concerned are expressed quantitatively.

Three groups of variables are normally recorded in crop experiments. These are:

1. Treatments, such as fertilizer rates, varieties, and weed control methods, which are generated from one or more management practices and are the primary focus of the experiment.
2. Environmental factors, such as rainfall and solar radiation, which represent the portion of the environment that is not within the researcher's control.
3. Responses, which represent the biological and physical features of the experimental units that are expected to be affected by the treatments being tested.

Response to treatments can be exhibited either by the crop, in terms of changes in such biological features as grain yield and plant height (to be called *crop response*), or by the surrounding environment in terms of changes in such features as insect incidence in an entomological trial and soil nutrient in a fertility trial (to be called *non-crop response*). Because agricultural research focuses primarily on the behavior of biological environmental factors, and responses that are usually evaluated in crop research are:

1. Association between Response Variables. Crop performance is a product of several crop and non-crop characters. Each, in turn, is affected by the treatments. All these characters are usually measured simultaneously, and their association with each other can provide useful information about how the treatments influenced crop response. For example, in a trial to determine the effect of plant density on rice yield, the association between yield and its components, such as number of tillers or panicle weight, is a good indicator of the indirect effect of treatments; grain yield is increased as a result of increased tiller numbers, or larger panicle size, or a combination of the two.

Another example is in a varietal improvement program designed to produce rice varieties with both high yield and high protein content. A positive association between the two characters would indicate that varieties with both high yield and high protein content are easy to find, whereas a negative association would indicate the low frequency of desirable varieties.

2. Association between Response and Treatment. When the treatments are quantitative, such as kilograms of nitrogen applied per hectare and numbers of plants per m², it is possible to describe the association between treatment and response. By characterizing such an association, the relationship between treatment and response is specified not only for the treatment levels actually tested but for all other intermediate points within the range of the treatments tested. For example, in a fertilizer trial designed to evaluate crop yield at 0, 30, 60, and 90 kg N/ha, the relationship between yield and nitrogen rate specifies the yields that can be obtained not only for the four nitrogen rates actually tested but also for all other rates of application between zero and 90 kg N/ha.

3. Association between Response and Environment. For a new crop management practice to be acceptable, its superiority must hold over diverse environments. Thus, agricultural experiments are usually repeated in different areas or in different crop seasons and years. In such experiments, association between the environmental factors (sunshine, rainfall, temperature, soil nutrients) and the crop response is important.

Regression analysis describes the effect of one or more variables (designated as *independent variables*) on a single variable (designated as the *dependent variable*) by expressing the latter as a function of the former. For this analysis, it is important to clearly distinguish between the dependent and independent variable, a distinction that is not always obvious. For instance, in experiments on yield response to nitrogen, yield is obviously the dependent variable and nitrogen rate is the independent variable. On the other hand, in the example on grain yield and protein content, identification of variables is not obvious. Generally, however, the character of major importance, say grain yield, becomes the dependent variable and the factors or characters that influence grain yield become the independent variables.

Correlation analysis, on the other hand, provides a measure of the degree of association between the variables or the goodness of fit of a prescribed relationship to the data at hand.

Regression and correlation procedures can be classified according to the number of variables involved and the form of the functional relationship between the dependent variable and the independent variables. The procedure is termed *simple* if only two variables (one dependent and one independent variable) are involved and *multiple*, otherwise. The procedure is termed linear if the form of the underlying relationship is linear and *nonlinear*, otherwise. Thus, regression and correlation analysis can be classified into four types:

1. simple linear regression and correlation
2. multiple linear regression and correlation
3. simple nonlinear regression and correlation
4. multiple nonlinear regression and correlation

The relationship between any two variables is linear if the change is constant throughout the whole range under consideration. The graphical representation of a linear relationship is a straight line, as illustrated in graph (a) below.

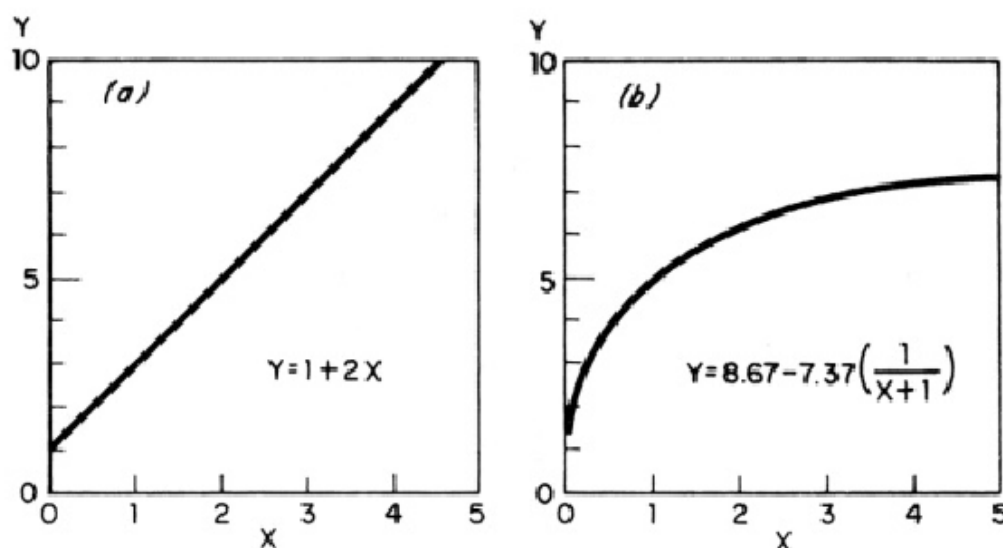


Illustration of a linear (a), and a nonlinear (b), relationship between the dependent variable Y and the independent variable X.

In graph (a), Y constantly increases two units for each unit change in X throughout the whole range of X values from 0 to 5: Y increases from 1 to 3 as X changes from 0 to 1, and Y increases from 3 to 5 as X changes from 1 to 2, and so on.

In graph (b), the pattern is not linear, rather, it which usually starts slowly, increases to a fast rate at intermediate growth stages, and slows toward the end of the cycle. This is an example of nonlinear relationship.

The functional form of the linear relationship between a dependent variable Y and an independent variable X is represented by the equation:

$$Y = \alpha + \beta X$$

where

α is the intercept of the line on the Y axis and

β , the linear regression coefficient, is the slope of the line or the amount of change in Y for each unit change in X .

With the two parameters of the linear relationship specified, the value of the dependent variable Y , corresponding to a given value of the independent variable X within the range of X values considered, can be immediately determined by replacing X in the equation with the desired value and computing for Y .

When there is more than one independent variable, say k independent variables ($X_1, X_2, \dots, X_k$), the simple linear functional form of the equation $Y = \alpha + \beta X$ can be extended to the multiple linear functional form of where α is the intercept (value of Y when all X 's are zeroes) and β_i ($i = 1, \dots, k$), the partial regression coefficient associated with the independent variable X_i , represents the amount of change in Y for each unit change in X_i . Thus, in the multiple linear functional form with k independent variables, there are $(k + 1)$ parameters (i.e., $\beta_1, \beta_2, \dots, \beta_k$) that need to be estimated.

Simple Linear Regression and Correlation

For the simple linear regression analysis to be applicable, the following conditions must hold:

- There is only one independent variable X affecting the dependent variable Y .
- The relationship between Y and X is known, or can be assumed, to be linear. Although these two conditions may seem too restrictive, they are often satisfied for data from controlled experiments.

Most controlled experiments are designed to keep the many factors that can simultaneously influence the dependent variable constant, and to vary only the factor (treatment) being investigated. In a nitrogen fertilizer trial, for example, all other factors that can affect yield, such as phosphorus application, potassium application, plant population, variety, and weed control, are carefully controlled throughout the experiment. Only nitrogen rate is varied. Consequently, the assumption that the rate of nitrogen application is the major determinant of the variation in the yield data is satisfied.

In contrast, if data on grain yield and the corresponding rate of nitrogen application were collected from an experiment where other production factors were allowed to vary, the assumption of one independent variable would not be satisfied and, consequently, the use of a simple regression analysis would be inappropriate.

The simple linear regression analysis deals with the estimation and tests of significance concerning the two parameters α and β in the equation $Y = \alpha + \beta X$. It should be noted that because the simple linear regression analysis is performed under the assumption that there is a linear relationship between X and Y , it does not provide any test as to whether the *best* functional relationship between X and Y is indeed linear.

The data required for the application of the simple linear regression analysis are the n pairs (with $n > 2$) of Y and X values. As an example, consider a study of nitrogen response using data from a fertilizer trial involving 11 nitrogen rates, the 11 pairs of Y and X values would be the 11 pairs of mean yield (Y) and nitrogen rate (X).

Sample	Nitrogen Level in kg (X)	Yield (Y)
1	0	4.2
2	5	10.0
3	10	8.8
4	15	16.6
5	20	14.4
6	25	22.5
7	30	18.4
8	35	26.4
9	40	27.4
10	45	28.3
11	50	32.4
Total	275	209.4
Mean	25.0	19.0

The following computations are needed:

Sum of Squares of X (designated as Σx^2)

$$\Sigma x^2 = \Sigma (X)^2 - \frac{(\Sigma X)^2}{n} = (0)^2 + (5)^2 + \dots + (50)^2 - \frac{(275)^2}{11} = \boxed{2,750}$$

Sum of Squares of Y (designated as Σy^2)

$$\Sigma y^2 = \Sigma (Y)^2 - \frac{(\Sigma Y)^2}{n} = (4.2)^2 + (10.0)^2 + \dots + (32.4)^2 - \frac{(209.4)^2}{11} = \boxed{834.97}$$

Sum of Cross-Products of X and Y (designated as Σxy)

$$\Sigma xy = \Sigma (XY) - \frac{(\Sigma X \Sigma Y)}{n} = (0 \times 4.2) + (5 \times 10.0) + \dots + (50 \times 32.4) - \frac{275 \times 209.4}{11} = \boxed{1,468}$$

Regression coefficient (b)

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{1,468}{2,750} = \boxed{0.53}$$

Intercept (a)

$$a = \bar{Y} - b\bar{X} = 19.0 + 0.53(25.0) = \boxed{5.69}$$

The regression equation now is:

$$\boxed{Y = 5.69 + 0.53X}$$

The equation indicates that the value of dependent variable Y is 5.69 when the value of independent variable X is zero and the value of Y changes at the rate of 0.53 unit for every unit change in the value of X .

With the regression equation computed, the estimated value of dependent variable Y (designated as Y') can now be computed.

$$Y1' = 5.69 + 0.3(0) = \boxed{5.7}$$

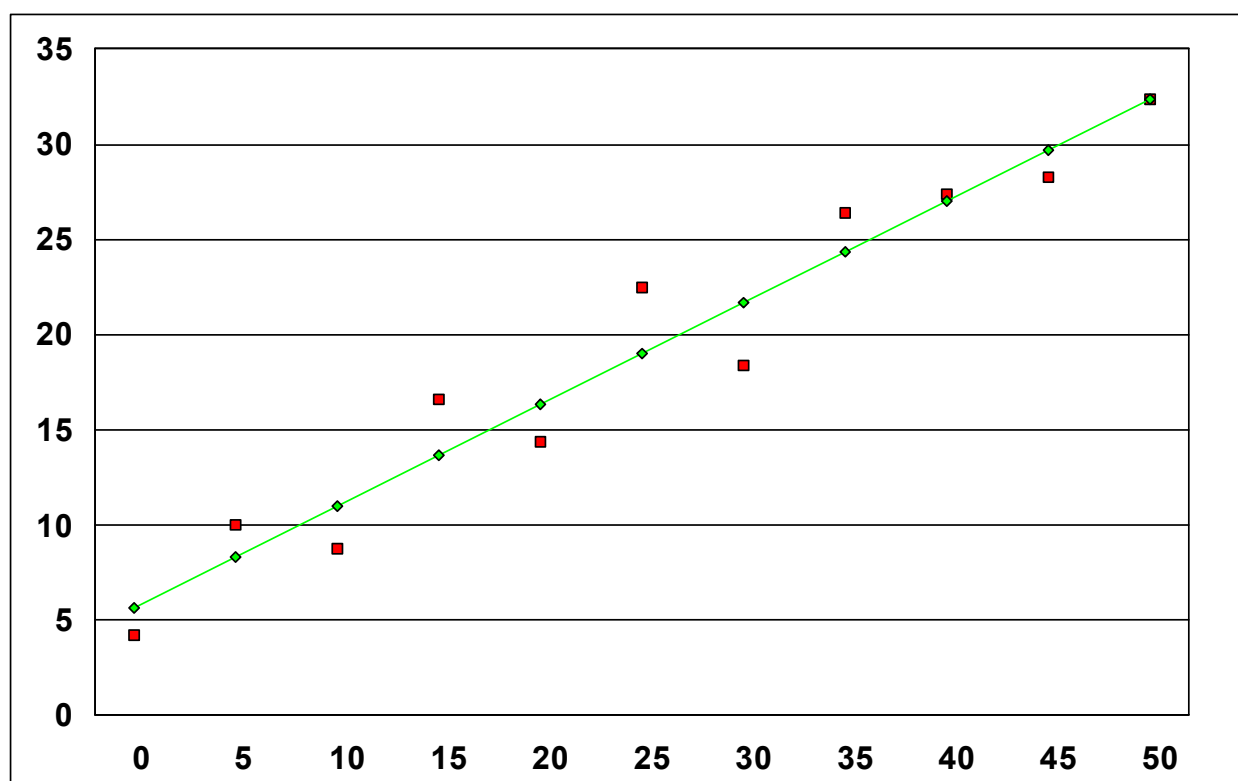
$$Y2' = 5.69 + 0.53(5) = \boxed{8.4}$$

$$Y11' = 5.69 + 0.53(50) = \boxed{32.4}$$

The values of the estimated yield are shown in the next table.

Sample	Nitrogen level in kg (X)	Yield (Y)	Estimated Yield (Y')
1	0	4.2	5.7
2	5	10.0	8.4
3	10	8.8	11.0
4	15	16.6	13.7
5	20	14.4	16.4
6	25	22.5	19.0
7	30	18.4	21.7
8	35	26.4	24.4
9	40	27.4	27.0
10	45	28.3	29.7
11	50	32.4	32.4
Mean	25.0	19.0	19.0

A graphic presentation is shown below:



The square figures represent the observed values of the dependent variable X and the line represents the regression line as computed by the equation $\underline{Y = 5.69 + 0.534X}$. Note that in the example, the original values do not coincide with the corresponding points in the regression line.

The difference between the original observation and the computed value becomes the residual error which is used to test the significance of **b** and **a**.

After determining the regression equation, test of significance is needed to determine if the linear response (**b**) is significant. First, compute for residual mean square ($s^2_{y.x}$).

$$s^2_{y.x} = \frac{\Sigma y^2 - \frac{(\Sigma xy)^2}{\Sigma x^2}}{n - 2} = \frac{834.97 - \frac{(1,468)^2}{2,750}}{11 - 2} = \boxed{4.6655}$$

Compute the t-test for **b** as:

$$t_b = \frac{b}{\sqrt{\frac{s^2_{y.x}}{\Sigma x^2}}} = \frac{0.534}{\sqrt{\frac{4.6655}{2,750}}} = \boxed{12.960^*}$$

The t_b value is compared with the t-table value at 5% level and n-2 df (= 2.262). Since t_b value is greater than t-table value, the regression equation is considered significant, and that the yield values follow a linear response.

The (100 – α)% confidence interval for β is constructed:

$$C.I. = b \pm t_\alpha \sqrt{\frac{s^2_{y.x}}{\Sigma x^2}} = 0.534 \pm 2.262 \sqrt{\frac{4.6655}{2,750}} = 0.534 \pm 0.093 = 0.441 \text{ and } 0.626$$

The C.I. values indicate that the increase in yield for every 1 unit increase in the rate of N applied, within the range of 0 to 50 kg, is expected to fall between 0.44 and 0.63 unit of yield, 95% of the time.

Simple Linear Correlation Analysis

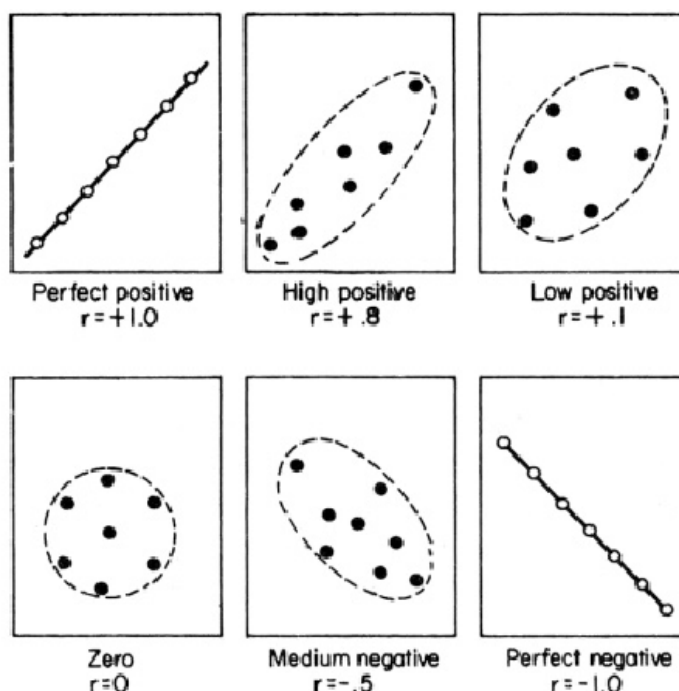
The simple linear correlation analysis deals with the estimation and test of significance of the simple linear correlation coefficient r , which is a measure of the degree of linear association between two variables X and Y . Computation of the simple linear correlation coefficient is based on the amount of variability in one variable that can be explained by a linear function of the other variable.

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}}$$

The result is the same whether Y is expressed as a linear function of X , or X is expressed as a linear function of Y . Thus, in the computation of the simple linear correlation coefficient, there is no need to specify which variable is the cause and which is the consequence, or to distinctly differentiate between the dependent and the independent variable, as is required in the regression analysis.

The value of r lies within the range of –1 and +1, with the extreme value indicating perfect linear association and the mid-value of zero indicating no linear association between the two variables. An intermediate value of r indicates the portion of variation in one variable that can be accounted for by the linear function of the other variable. For example, with an r value of 0.8, the implication is that 64% $[(100)(r^2) = (100)(0.8)^2 = 64\%]$ of the variation in the variable Y can be explained by the linear function of the variable X .

The minus or plus sign attached to the r value indicates the direction of change in one variable relative to the change in the other. That is, the value of r is negative when a positive change in one variable is associated with a negative change in another, and positive when the two variables change in the same direction.



Graphical representations of various values of simple correlation coefficient r .

In agricultural research, there are two common applications of the simple linear correlation analysis:

- It is used to measure the degree of association between two variables with a well-defined cause and effect relationship that can be defined by the linear regression equation $Y = \alpha + \beta X$.
- It is used to measure the degree of linear association between two variables in which there is no clear-cut cause and effect relationship.

A note of caution is added here concerning the magnitude of the computed r value and its corresponding degree of freedom. The tabular r values decrease sharply with the increase in the degree of freedom, which is a function of n (i.e., the number of pairs of observations used in the computation of the r value). Thus, the smaller n is, the larger the computed r value must be to be declared significant. With $n = 4$, the seemingly high value of the computed r of 0.985 is still not significant at the 1% level. On the other hand, with $n = 9$, a computed r value of 0.8 would have been declared significant at the 1% level. Thus, the practical importance of the significance and the size of the r value must be judged in relation to the sample size n . It is, therefore, a good practice to always specify n in the presentation of the regression and correlation result

Correlation between response and treatment

Using the data in the example in regression analysis, the procedures for the estimation and test of significance of a simple linear correlation coefficient between two variables X and Y are:

1. Compute the means X and Y , corrected sums of squares of X and Y , and corrected sum of cross products of X and Y . This was done already in computing for regression.

$$\begin{aligned} SS_X &= 2,750 \\ SS_Y &= 834.97 \\ SCP_{XY} &= 1,468 \end{aligned}$$

2. Compute simple linear regression as:

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{1,468}{\sqrt{(2,750)(834.97)}} = \boxed{0.969}$$

3. Test the significance of the simple linear correlation coefficient by comparing the computed r value to the tabular r value with $(n - 2)$ degrees of freedom. The simple linear correlation coefficient is declared significant at the α level of significance if the absolute value of the computed r is greater than the corresponding tabular r value at the α level of significance. The r tabular values are

$$\alpha_{0.05, 9 \text{ df}} = \underline{0.602}$$

$$\alpha_{0.01, 9 \text{ df}} = \underline{0.732}.$$

Since the r value of 0.969 is greater than the tabular value at 1% level, the rates of nitrogen and yield are very highly correlated. The computed r value of 0.969 indicates that 97% $[(100)(0.969)^2]$ of the variation in the mean yield is accounted for by the linear function of the rate of nitrogen applied. The relatively high r value obtained is also indicative of the closeness between the estimated regression line and the observed points. Within the range of 0 to 50 kg N/ha, the linear relationship between mean yield and rate of nitrogen applied seems to fit the data adequately.

Correlation between two responses

Consider a trial of 10 rice varieties with data on yield and protein content. Since the treatment is about 10 varieties, it is not possible to make a correlation between the treatment and any of the responses. But the degree of association between the 2 responses can be determined. In this case, it is not clear whether there is a cause and effect relationship between the two variables and, even if there were one, it would be difficult to specify which is the cause and which is the effect. Hence, the simple linear correlation analysis is applied to measure the degree of linear association between the two variables without specifying the causal relationship.

Variety	Protein (%)	Yield
1	1.61	15.86
2	1.82	10.45
3	1.71	14.24
4	1.78	9.98
5	2.02	11.45
6	1.87	13.46
7	1.98	6.48
8	1.32	20.84
9	1.91	11.60
10	1.40	21.30
Mean	1.74	13.57

Following the procedures in the previous example, the following values are obtained:

$$SS_X = \mathbf{0.49956}$$

$$SS_Y = \mathbf{199.57664}$$

$$SCP_{XY} = \mathbf{-9.03692}$$

And correlation is computed as:

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{-9.03692}{\sqrt{(0.49956)(199.57664)}} = \boxed{-0.905}$$

Because the computed r value exceeds both tabular r values, we conclude that the simple linear correlation coefficient is significantly different from zero at the 1% probability level. This significant, high $-r$ value indicates that there is strong evidence that the soluble protein nitrogen and the total yield in the different rice varieties are highly associated with one another in a linear way: varieties with high soluble protein nitrogen have a lower yield, or varieties with high yield have a low protein content.

Multiple Linear Regression and Correlation

The simple linear regression and correlation analysis has one major limitation. That is, that it is applicable only to cases with one independent variable. However, with the increasingly accepted perception of the interdependence between factors of production and with the increasing availability of experimental procedures that can simultaneously evaluate several factors, researchers are increasing the use of factorial experiments. Thus, there is a corresponding increase in need for use of regression procedures that can simultaneously handle several independent variables.

Regression analysis involving more than one independent variable is called *multiple regression analysis*. When all independent variables are assumed to affect the dependent variable in a linear fashion and independently of one another, the procedure is called *multiple linear regression analysis*. A multiple linear regression is said to be operating if the relationship of the dependent variable Y to the k independent variables $X_1, X_2, \dots, X_k$ can be expressed as

$$Y = a + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

The procedure is illustrated for a case where $k = 2$, using the data on grain yield (Y), plant height (X_1), and tiller number (X_2).

Sample	Plant Height (X_1)	Tiller Number (X_2)	Grain Yield (Y)
1	0	24	4.2
2	5	35	10.0
3	10	32	8.8
4	15	25	16.6
5	20	28	14.4
6	25	36	22.5
7	30	39	18.4
8	35	45	26.4
9	40	48	27.4
10	45	65	28.3
11	50	69	32.4
Total	275	446	209.4
Mean	25.0	40.6	19.0

With $k = 2$, the multiple linear regression equation is expressed as:

$$Y = a + \beta_1 X_1 + \beta_2 X_2$$

The steps are:

1. Compute the mean and the corrected sum of squares for each of the $(k + 1)$ variables $Y, X_1, X_2, \dots, X_k$, and the corrected sums of cross-products for all possible pair-combinations of the $(k + 1)$ variables.

Sum of Squares of Y (designated as ΣY^2)

$$\Sigma Y^2 = \Sigma(Y)^2 - \frac{(\Sigma Y)^2}{n} = (4.2)^2 + (10.0)^2 + \dots + (32.4)^2 - \frac{209.4^2}{11} = \boxed{834.97}$$

Sum of Squares of X_1 (designated as ΣX_1^2)

$$\Sigma X_1^2 = \Sigma(X_1)^2 - \frac{(\Sigma X_1)^2}{n} = (0)^2 + (5)^2 + \dots + (50)^2 - \frac{275^2}{11} = \boxed{2,750}$$

Sum of Squares of X_2 (designated as ΣX_2^2)

$$\Sigma X_2^2 = \Sigma(X_2)^2 - \frac{(\Sigma X_2)^2}{n} = (24)^2 + (35)^2 + \dots + (69)^2 - \frac{446^2}{11} = \boxed{2,282.7}$$

Sum of Cross-Products of X_1 and X_2 (designated as $\Sigma X_1 X_2$)

$$\Sigma X_1 X_2 = \Sigma(X_1 X_2) - \frac{(\Sigma X_1 X_2)}{n} = (0 \times 24) + (5 \times 35) + \dots + (50 \times 69) - \frac{(275 \times 446)}{11} = \boxed{2,220.0}$$

Sum of Cross-Products of X_1 and Y (designated as $\Sigma X_1 Y$)

$$\Sigma X_1 Y = \Sigma(X_1 Y) - \frac{(\Sigma X_1 Y)}{n} = (0 \times 4.2) + (5 \times 10.0) + \dots + (50 \times 32.4) - \frac{(275 \times 209.4)}{11} = \boxed{1,468.0}$$

Sum of Cross-Products of X_2 and Y (designated as $\Sigma X_2 Y$)

$$\Sigma X_2 Y = \Sigma(X_2 Y) - \frac{(\Sigma X_2 Y)}{n} = (24 \times 4.2) + (35 \times 10.0) + \dots + (69 \times 32.4) - \frac{(446 \times 209.4)}{11} = \boxed{1,166.3}$$

2. Compute for the 2 regression coefficients (b_1 and b_2). Although the derivation will not be shown here, the formula for the 2 regression coefficients are:

$$b_1 = \frac{(\Sigma X_2^2)(\Sigma X_1 Y) - (\Sigma X_1 X_2)(\Sigma X_2 Y)}{(\Sigma X_1^2)(\Sigma X_2^2) - (\Sigma X_1 X_2)^2} = \frac{(2,282.7)(1,468) - (2,220)(1,166.3)}{(2,750)(2,282.7) - (2,220)^2} = \boxed{0.565}$$

$$b_2 = \frac{(\Sigma X_1^2)(\Sigma X_2 Y) - (\Sigma X_1 X_2)(\Sigma X_1 Y)}{(\Sigma X_1^2)(\Sigma X_2^2) - (\Sigma X_1 X_2)^2} = \frac{(2,750)(1,166.3) - (2,220)(1,468)}{(2,750)(2,282.7) - (2,220)^2} = \boxed{-0.038}$$

3. Compute for intercept (a)

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 = 19.0 + 0.565(25.0) + (-0.038)(40.6) = \boxed{6.471}$$

The regression equation relating to plant height (X_1) and tiller number (X_2) to yield (Y) is:

$$Y = 6.471 + 0.565X_1 - 0.038X_2$$

4. Compute the following:

a. Sum of squares due to regression (SSR), as:

$$SSR = \sum(b_i)(\sum x_i y) = (0.564)(1,4680) + (-0.038)(1,166.3) = 784.365$$

b. Residual sum of squares (SSE), as:

$$SSE = \sum y^2 - SSR = 834.97 - 784.365 = 50.600$$

c. Coefficient of determination (R^2), as:

$$R^2 = \frac{SSR}{\sum y^2} = \frac{784.365}{834.97} = 0.939$$

The coefficient of determination R^2 measures the contribution of the linear function of k independent variables to the variation in Y . It is usually expressed in percentage. Its square root (i.e., R) is referred to as the *multiple correlation coefficient*.

Thus, 93.9% of the total variation in the yields of eight rice varieties can be accounted for by a linear function, involving plant height and tiller number.

5. Test the significance of R^2 by computing the **F-value**

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{(n - k - 1)}} = \frac{\frac{784.365}{2}}{\frac{50.600}{(11 - 2 - 1)}} = 62.00$$

The computed F value is compared to the tabular F values with $f_1 = k$ and $f_2 = (n - k - 1)$ degrees of freedom. The coefficient of determination R^2 is said to be significant (significantly different from zero) if the computed F value is greater than the corresponding tabular F value at the prescribed level of significance.

The tabular F values with $f_1 = 2$ and $f_2 = 8$ degrees of freedom are:

4.46 at the 5% level of significance, and

18.86 at the 1% level.

Because the computed F value is larger than the tabular F value at the 1% level, the estimated multiple linear regression $Y = 6.471 + 0.565(X_1) - 0.038(X_2)$ is highly significant at the 1% level of significance. Thus, the combined linear effects of plant height and tiller number contribute highly significantly to the variation in yield.

More than 2 independent variables

It is possible to compute linear regression and correlation coefficients involving more than 2 independent variables, however, the derivation of formula for the different regression coefficients ($b_1, b_2, \dots, b_k$) becomes very complicated so it is not practical to do it manually.

A technique using abbreviated Doolittle method may be used when dealing with more than 2 independent variables, however, this will not be discussed in this hand-out due to its complexity. At present, computer programs are readily available to do the complicated and tedious computations.

CHI-SQUARE TEST

The chi-square test is most commonly used to test hypotheses concerning the frequency distribution of one or more populations. We focus on three uses of the chi-square test that are most common in agricultural research: analysis of attribute data, test for homogeneity of variance, and test for goodness of fit.

Analysis of Attribute Data

Data from an agricultural experiment can either be measurement data or attribute data. Measurement data is specified along a continuous numerical scale, such as yield, plant height, and protein content: but attribute data is concerned with a finite number of discrete classes. The most common types of attribute data are those having two classes, which consist of the presence or absence of an attribute such as male or female, success or failure, effective or ineffective, and dead or alive. Examples of attribute data with more than two classes are varietal classification, color classification, and tenure status of farmers.

The number of discrete classes in attribute data may be specified based on one or more classification criteria. When only one criterion is used, attribute data is referred to as a one-way classification. Presence or absence of one character, color classification of a plant tissue, and tenure status of farmers are illustrations of attribute data with one-way classification. When more than one classification criterion is used to specify the classes in attribute data, such data may be referred to as a two-way classification, a three-way classification, and so on, depending on the number of classification criteria used. Attribute data with two-way classification form an $r \times c$ two-way classification, or an $r \times c$ contingency table, where r and c denote the number of classes in the two classification criteria used. For example, if rice varieties in a variety trial are classified based on two criteria—color of its leaf blade (green or purple) and varietal type (indica, japonica, or hybrid)—the resulting attribute data represent a 2×3 contingency table. Note that the contingency table progresses to three-way, four-way, and so on, with successive additions of more classification criteria.

In general, attribute data are obtained when it is not possible to use measurement data. However, in some special cases experimental materials may be classified into discrete classes despite the availability of a quantitative measurement. For example, plants can be classified into three discrete height classes (tall, intermediate, or short) instead of being measured in centimeters. Or, vertical resistance to an insect pest may be scored on a scale from 0 through 9 instead of measuring the actual percentage of plant damage or of insect incidence.

There are three important applications of the chi-square test in the analysis of attribute data:

- Test for a fixed-ratio hypothesis
- Test for independence in a contingency table
- Test for homogeneity of ratio

Test for a Fixed-Ratio Hypothesis

As the name implies, the chi-square test for a fixed-ratio hypothesis is a technique for deciding whether a set of attribute data conforms to a hypothesized frequency distribution that is specified on the basis of some biological phenomenon.

a. With two classes:

$$\chi^2 = \frac{(|n_1 - E_1| - 0.5)^2}{E_1} + \frac{(|n_2 - E_2| - 0.5)^2}{E_2}$$

where $||$ refers to absolute value; and n_1 and n_2 are observed values and E_1 and E_2 are the expected values, as defined previously, of class 1 and class 2, respectively.

As an example, observations were made on the preferred growing medium of a particular bacterium to determine if the preference is the same for PDA and coconut water. The observations showed 128 bacterial colonies in PDA and 172 in coconut water. Is the hypothesis of equal preference valid or not?

With $r_1 = r_2 = 1$, $n_1 = 128$, and $n_2 = 172$, the values of E_1 and E_2 are computed following the formula as:

$$E_1 = E_2 = \frac{(1)(n_1 + n_2)}{1 + 1} = \frac{128 + 172}{2} = \boxed{150}$$

and chi-square is:

$$\chi^2 = \frac{(|128 - 150| - 0.5)^2}{150} + \frac{(|172 - 150| - 0.5)^2}{150} = \frac{(21.5)^2}{150} + \frac{(21.5)^2}{150} = \boxed{6.16}$$

Compare the computed χ^2 value with the χ^2 at $(n - 1)$ df which is 6.63 for 1% and 3.84 for 5%. Since computed value is smaller than 1% tabular value but larger than 5% value, the χ^2 test showed significant difference between the two observations. It means that the proposed equal growing preference is not true and that the bacteria prefer coconut water over PDA.

b. With more than two classes:

$$\chi^2 = \sum_{i=1}^p \frac{(n_i - E_i)^2}{E_i}$$

where p is the number of classes, n_i is the observed number of units falling into class i , and E_i is the number of units expected to fall into class i assuming that the hypothesized ratio holds. E_i is computed as:

$$E_i = \frac{(r_i) \left(\sum_{i=1}^p n_i \right)}{\sum_{i=1}^p r_i}$$

where $r_1 : r_2 : \dots : r_p$ is the hypothesized ratio.

As an example, a plant breeder is studying a cross between a sweet maize inbred line with yellow kernels and a flint maize inbred line with white kernels. He would like to know whether the ratio of kernel type and color in the F_2 population follows the normal di-hybrid ratio of 9:3:3:1. From the F_1 plants produced by crossing the two inbred lines, he obtains F_2 kernels and classifies them into four categories according to kernel color (yellow or white) and kernel type (flint or sweet) as follows: yellow flint (YF), yellow sweet (YS), white flint (WF), and white sweet (WS). Suppose he examines 800 F_2 kernels and finds that 496 are yellow flint, 158 are yellow sweet, 112 are white flint, and the rest (34) are white sweet. He then asks: does the observed ratio of 496:158:112:34 deviate significantly from the hypothesized ratio of 9:3:3:1?

Compute for the expected frequency of each classification based on the hypothesized ratio. The sum of the hypothesized ratio is: $9+3+3+1 = 16$

$$E_{YF} = \frac{800 \times 9}{16} = \boxed{450}$$

$$E_{YS} = \frac{800 \times 3}{16} = \boxed{150}$$

$$E_{WF} = \frac{800 \times 3}{16} = \boxed{150}$$

$$E_{WS} = \frac{800 \times 1}{16} = \boxed{50}$$

The X^2 is computed as:

$$X^2 = \frac{(496 - 450)^2}{450} + \frac{(158 - 150)^2}{150} + \frac{(112 - 150)^2}{150} + \frac{(34 - 50)^2}{50}$$

$$= 4.70 + 0.43 + 9.63 + 5.12 = \boxed{19.88}$$

The computed X^2 value is compared with the tabular value at $(n - 1)$ df which is 11.34 for 1% and 7.81 for 5%. Since the computed X^2 value is larger than the tabular values, the hypothesis that the observed ratio follows the expected ratio is not true. It means more than two genes may be involved in the expression of the two traits.

Test for Independence in a Contingency Table

One of the most common use of X^2 test is in a two-way classification often used when conducting survey. A classical example is adoption of a particular technology by a group of respondents that can be divided into discrete classes. As an example, consider a group of farmers being interviewed about their response to the introduction of a new wheat variety. The farmers can be classified into three categories: owner-operator, share-rent farmer, and fixed-rent farmer. The response can be classified into two classes: adopter and non-adopter. The two-way classification with the hypothetical data becomes:

Tenure status	Adopter	Semi-adopter	Non-adopter	Row Total
Owner operator	204	110	52	256
Share-rent farmer	84	54	20	104
Fixed-rent farmer	12	8	6	18
Column Total	300	172	78	378

The expected frequency is computed based on the column, row and grand totals. For Owner operator adopter, the expected frequency is:

$$\frac{(300 \times 256)}{378} = \boxed{203.2}$$

The other values can be computed using the same formula. The values are shown in the next table:

Tenure status	Adopter	Semi-adopter	Non-adopter
Owner operator	203.2	116.5	52.8
Share-rent farmer	82.5	47.3	21.5
Fixed-rent farmer	14.3	8.2	3.7

The X^2 is computed as:

$$\frac{(204 - 203.2)^2}{203.2} + \frac{(110 - 116.5)^2}{116.5} + \dots + \frac{(6 - 3.7)^2}{8.2} = \boxed{3.22}$$

The computed X^2 is compared with the tabular value at $(c-1)(r-1)$ df ($2 \times 2 = 4$) which is 13.28 for 1% and 9.49 for 5%. Since the computed X^2 value is smaller than the 5% tabular value, the hypothesis that the adoption of the new wheat variety is independent with regards to type of farmer is accepted.

Test for Homogeneity of Variance

One of the assumptions in a valid analysis of variance is homogeneity (or uniformity) or variances of the different treatments or trials. If the range of variances is too large, the analysis is not considered valid. One of the tests used to determine this is Bartlett's test which basically uses chi-square. Consider an example in Combined Analysis of 4 variety trials where the respective error mean squares and their log values are:

	EMS	log(EMS)
Trial 1 =	4.6539	0.6678
Trial 2 =	4.7242	0.6743
Trial 3 =	5.0896	0.7067
Trial 4 =	26.0994	1.4166
TOTAL	40.5671	3.4655

Compute the pooled estimate of variance as:

$$s_p^2 = \frac{\sum s_i^2}{k} = \frac{40.5671}{4} = \boxed{10.1418}$$

Chi-square is computed as:

$$X^2 = \frac{(2.3026)(f)(k \log s_p^2 - \sum \log s_i^2)}{1 + \frac{k+1}{3 \times k \times f}}$$

where 2.3026 = constant

f = error degree of freedom of each trial = 29

k = number of variances = 4

$$X^2 = \frac{(2.3026)(21)(4 \times \log(10.1418) - 3.4655)}{1 + \frac{4+1}{3 \times 4 \times 21}} = \frac{48.3546 \times 0.5590}{1.0198413} = \frac{27.0300}{1.0198413} = \boxed{26.50}$$

The computed X^2 is compared with the tabular value at $(k-1)$ df ($4-1=3$) which is 11.34 for 1% and 7.81 for 5%. Since the computed X^2 value is larger than both 1% and 5% tabular value, the hypothesis that the variances are homogenous is rejected. The decision to exclude the trial with the highest variance in the combined analysis is valid.

SUMMARIZED COMPARISON OF THE THREE EXTENSIONS OF THE BASIC DESIGNS

ITEM	FACTORIAL	SPLIT-PLOT	STRIP-PLOT
Application	For experiments where the main factors A & B and the interaction AxB are of equal importance	One factor is more important than the other such that a more precise detection of significance is desired. The more important factor as assigned as sub-plot and the lesser important factor as the main plot. One factor requires larger plot to prevent or minimize border effects, or for easier management.	Factors A & b are equally important but interaction AxB is more important than either A or B. Both factors require larger plots to prevent or minimize border effects, or for easier management.
Randomization CRD	Factors x block combinations are randomized in the whole experimental area	Main plot (Factor A) is randomized in the whole experimental area. Sub-plot (Factor B) is randomized in each main plot.	Not applicable
RCBD	Factor combinations are randomized in each block independent of randomization of other blocks.	Main plot (Factor A) is randomized in each block independent of other blocks. Sub-plot (Factor B) is randomized in each main plot.	For each block, Factor A is randomized in one direction (e.g. vertical) while factor B is randomized in the opposite direction (e.g. horizontal). Randomization is independent on each block.
Number of error terms	Only one (1) error term	Number of error terms equals the number of main factors	Number of error terms equals the total number of main factors and interactions
Formula of error term CRD	Replication is nested within the factor combination	Replication is nested within the main factor (Factor A); succeeding error terms are the interaction between the replication and the main factor [nested in the factor(s) above it] immediately preceding it.	Not applicable
RCBD	Interaction between factor combinations and replication	Interaction between replication and the main factor [nested in the factor(s) above it] preceding them.	Interaction between the replication and the main factor or interaction immediately preceding it.
Test of significance	The single pooled error is used to test significance of all main factors and interactions	Error A is used to test significance of the main factor and block, succeeding error terms are used to test significance of the factor and interaction(s) preceding them.	Individual error terms is used to test significance of the main factor or interaction immediately preceding them.

TABLE OF RANDOM NUMBERS

09724	85125	48477	42783	70473	52491	09137	26453	02148	30742	57866	87037	57196	47916	15960	13036	84639	30186	48347	40780
14620	95430	12951	81953	17629	83603	66875	93650	91487	37190	61684	96598	28043	25325	81767	20792	39823	48749	79489	39329
56919	17803	85069	61594	92086	85437	92086	53045	13847	36207	96807	38025	12858	02821	41319	96808	80057	67617	04478	18501
97310	78209	51263	52396	82681	82611	74070	83468	47615	23721	00810	85323	18076	18689	94178	15959	11072	39823	63756	18501
07585	28040	26939	64531	70570	98412	70858	78195	18295	32585	06461	45073	88350	35246	15851	16129	57460	34512	10243	47635
25950	85189	69374	37904	06759	70799	09701	47920	75108	35023	82197	50260	84476	72259	95036	50223	37215	07692	76527	80764
82973	16405	81497	20863	94072	83615	95249	63461	46857	31924	47430	50260	84476	72259	95036	69176	21753	89713	07468	61522
60819	27364	59081	72635	49180	72537	46950	83643	84697	81736	25043	52002	03643	69512	71294	69095	96340	58341	06381	19219
74208	69516	79530	47649	53046	95420	39206	47315	53290	30853	34718	59598	96345	48646	06387	54221	40422	93251	43456	89176
59041	38475	03615	84093	49731	62748	41857	69420	79762	01935	23965	11667	09746	60791	47409	32406	80874	74010	91548	79394
48480	50075	11804	24956	72182	59649	28493	75091	82753	15040	67207	47166	44917	94177	31846	73872	92835	12596	64807	23978
39412	03642	87497	29735	14308	46309	70949	83538	53920	47192	08261	71627	96865	75380	42735	19446	78478	35681	07769	18230
95318	16249	49512	35408	21814	07564	16284	50969	94969	26081	10289	83203	14456	32978	82587	64377	54270	47869	66444	68728
72094	16385	90185	72635	86259	38352	25354	36853	15395	20405	75622	93145	14951	46603	84176	17564	53965	80771	10453	87972
63158	49753	84279	56496	30618	23973	94710	25237	48544	38405	62557	05584	08081	46603	84176	17564	53965	80771	10453	87972
19082	73645	09182	73649	58823	95208	49635	01420	46768	45362	51695	70743	68481	57937	62634	86727	69563	29308	51729	10453
94252	77489	62434	20965	20247	04894	34739	57052	29950	54833	69596	25201	56536	54517	86909	92927	07827	28271	52075	52075
15232	84146	87729	65584	83641	19468	92764	19609	52095	74151	75284	36241	59749	81958	44318	28067	67638	72196	54648	36886
48392	06359	47040	05695	79799	05342	25989	85397	74372	18079	64082	68375	84991	48334	38970	82481	94725	56930	47276	27641
72020	18895	84948	53072	74573	19520	54212	21539	48207	95920	94649	33784	30361	32627	74667	48289	29629	61248	34939	76162
09394	59842	39573	51630	78548	06461	27837	24953	28316	42610	25261	28316	05692	82874	37083	73818	39897	97096	48508	81783
37950	77387	35495	48192	84518	30210	06566	21752	78967	45692	21967	90859	37178	34023	09397	55027	78758	51482	81867	26484
34800	28055	91570	99154	39603	76846	77183	50369	16501	68867	63749	41490	72232	71710	36489	15291	68579	83195	60186	78142
36435	75946	85712	06293	85621	97764	05792	34796	06096	63487	42869	78641	50359	80895	78641	50359	20497	91381	72319	83280
28187	31824	52265	80494	66428	15703	53126	53376	54205	91590	91729	08960	70364	14262	76861	06406	85253	57490	80497	54272
13838	79940	97007	67511	87939	68417	21786	09822	67510	23817	29806	57594	41320	29806	57594	41320	50929	18752	12856	09587
72201	08423	41489	15498	94911	79392	65362	19672	93682	84190	27650	57930	25216	67180	42352	41671	39787	90508	42479	60463
63435	45192	62020	47358	32286	41659	06374	47269	93682	62972	68318	14891	96592	44278	80631	82547	78178	97394	98513	29634
62367	45627	58317	76928	50274	28705	31842	96484	70904	23749	91423	83067	14837	03817	21850	39732	18603	27174	71319	82016
59038	96983	49218	57179	08062	25074	45060	50903	66578	41465	54574	54648	29265	63051	07586	78418	48489	05425	27931	84965
71254	81686	85861	63973	96086	89681	50212	92829	27698	62284	93987	91493	61816	09628	31397	17607	97095	47154	40798	06217
07896	62924	35682	42820	43646	37385	37236	72876	02323	77975	59854	13847	37190	47369	39657	45179	06178	58918	37965	32031
54614	19092	83860	11351	32533	56032	42009	16496	51396	95237	12636	80651	34352	52548	61636	97294	95394	71846	98148	47548
71433	54331	58437	03542	76797	50437	13576	49745	90804	80128	04856	51498	35242	60595	57125	24634	56276	30294	62698	12839
30176	71248	37983	06073	89096	43498	95782	70452	14651	12042	92417	96727	90734	84549	04236	02520	29057	22102	18358	95938
79072	87795	23294	61602	62921	38385	69546	47104	72917	66273	95723	05695	64543	12870	17646	25542	91526	91395	46359	52952
75014	96754	67151	82741	24283	64276	78438	70757	40749	85183	97643	47916	46058	34375	76054	67823	07381	96863	65765	29368
37390	75846	74579	94606	54959	35310	31249	42920	46046	73432	14398	48258	56272	10835	09898	71563	82459	37210	72547	82546
24524	32751	28350	43090	79672	94672	07091	15101	95390	38083	38715	36409	47286	96537	29890	04837	44262	24974	82081	79587
26316	20378	16474	62438	42496	35191	49368	30074	93346	29425	14020	16902	52324	27208	99811	60503	13967	70562	55274	64785
61085	96937	02520	86801	30980	58479	34924	25101	87373	61560	70051	31424	26201	88098	31019	36195	23032	92648	74724	68292
45836	41086	41283	97460	51798	29852	47271	49341	94156	49341	56602	58040	48323	37857	99639	10700	98176	34642	43428	39068
92103	19679	16921	65924	12521	31724	60336	01968	15971	07963	69874	15653	70998	02969	42103	01069	68736	52765	23824	31235
10317	82592	63205	12528	24367	15817	87172	76830	02350	76394	35242	79841	46481	17365	84609	26357	60470	35212	51863	00401
39764	21251	41749	43789	70565	35496	12479	52021	41843	83489	20364	89248	58280	41596	87712	97928	45494	78356	72100	32949
83594	95692	51910	23202	93736	10817	01242	10724	27035	67562	16572	14877	42927	46635	09564	45334	63012	47305	27136	19428
08087	01753	01787	79608	53164	74978	51631	07525	72656	80854	74256	48391	02159	67282	00649	40382	43087	34506	53229	08383
57819	39689	32509	87540	38150	47872	14614	18427	06725	69326	04653	15507	78424	21981	46854	61980	10745	51573	12717	25524
96957	81060	28587	60905	67404	80450	21082	16074	61437	24961	46545	87214	80308	45190	95334	30474	29771	73924	82713	69487
48426	43513	82950	79838	45149	07143	73967	23723	06909	75375	32077	23074	14924	52685	51808	12428	50970	47019	21993	43350

Points for the Distribution of F[5% (light type) and 1% (bold face type)]

f ₂	f ₁ Degrees of freedom (for greater mean square)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161 4,062	200 4,999	216 5,403	225 5,425	230 5,764	234 5,859	237 5,928	239 5,961	241 6,022	242 6,056	243 6,082	244 6,106	245 6,142	246 6,169	248 6,208	249 6,234	250 6,261	251 6,286	252 6,302	253 6,323	253 6,334	254 6,352	245 6,361	254 6,366
2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.36	19.37 99.37	19.38 99.39	19.39 99.40	19.40 99.41	19.41 99.42	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.47 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50	19.50 99.50
3	10.13 34.12	9.55 30.82	9.28 29.48	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.54 26.14	8.53 26.12
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.29	4.78 10.15	4.74 10.05	4.70 9.96	4.68 9.89	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.96	4.06 7.87	4.03 7.79	4.00 7.72	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.94	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.00 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.86
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.71	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33	2.71 4.31
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.78	2.91 4.71	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 4.01	2.55 3.93	2.54 3.91
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60
12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.92 4.65	2.85 4.50	2.80 4.39	2.76 4.30	2.72 4.22	2.69 4.16	2.64 4.05	2.60 3.98	2.54 3.88	2.50 3.78	2.46 3.70	2.42 3.61	2.40 3.56	2.36 3.49	2.35 3.46	2.32 3.41	2.31 3.38	2.30 3.36
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10	2.63 4.02	2.60 3.96	2.55 3.85	2.51 3.78	2.46 3.67	2.42 3.59	2.38 3.51	2.34 3.42	2.32 3.37	2.28 3.30	2.26 3.27	2.24 3.21	2.22 3.18	2.21 3.16
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.03	2.96 4.69	2.85 4.46	2.77 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.56 3.86	2.53 3.80	2.48 3.70	2.44 3.62	2.39 3.51	2.35 3.43	2.31 3.34	2.27 3.26	2.24 3.21	2.21 3.14	2.19 3.11	2.16 3.06	2.14 3.02	2.13 3.00
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.70 4.14	2.64 4.00	2.59 3.89	2.55 3.80	2.51 3.73	2.48 3.67	2.43 3.56	2.39 3.48	2.33 3.36	2.29 3.29	2.25 3.20	2.21 3.12	2.18 3.07	2.15 3.00	2.12 2.97	2.10 2.92	2.08 2.89	2.07 2.87
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.45 3.61	2.42 3.55	2.37 3.45	2.33 3.37	2.28 3.25	2.24 3.18	2.20 3.10	2.16 3.01	2.13 2.96	2.09 2.93	2.07 2.86	2.04 2.80	2.02 2.77	2.01 2.75

f ₂	f ₁ Degrees of freedom (for greater mean square)																								f ₂
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	1.96
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.36	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92	1.8
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	19
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	20
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	21
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.42	2.38	2.36	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	22
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	23
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	24
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	25
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.19	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	26
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67	27
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93	2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10	
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65	28
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06	
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64	29
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62	30
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01	
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59	32
	7.15	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96	
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57	34
	7.44	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91	
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55	36
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87	
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53	38
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84	

f ₂	f ₁ Degrees of freedom (for greater mean square)																								f ₂
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
40	4.08 7.31	3.23 5.18	2.84 4.31	2.61 3.83	2.45 3.51	2.34 3.29	2.25 3.12	2.18 2.99	2.12 2.88	2.07 2.80	2.04 2.73	2.00 2.66	1.95 2.56	1.90 2.49	1.84 2.37	1.79 2.29	1.74 2.20	1.69 2.11	1.66 2.05	1.61 1.97	1.59 1.94	1.55 1.88	1.53 1.84	1.51 1.81	40
42	4.07 7.27	3.22 5.15	2.83 4.29	2.59 3.80	2.44 3.49	2.32 3.26	2.24 3.10	2.17 2.96	2.11 2.86	2.06 2.77	2.02 2.70	1.99 2.64	1.94 2.54	1.89 2.46	1.82 2.35	1.78 2.26	1.73 2.17	1.68 2.08	1.64 2.02	1.60 1.94	1.57 1.91	1.54 1.85	1.51 1.80	1.49 1.78	42
44	4.06 7.24	3.21 5.12	2.82 4.26	2.58 3.78	2.43 3.46	2.31 3.24	2.23 3.07	2.16 2.94	2.10 2.84	2.05 2.75	2.01 2.68	1.98 2.62	1.92 2.52	1.88 2.44	1.81 2.32	1.76 2.24	1.72 2.15	1.66 2.06	1.63 2.00	1.58 1.92	1.56 1.88	1.52 1.82	1.50 1.78	1.48 1.75	44
46	4.05 7.21	3.20 5.10	2.81 4.24	2.57 3.76	2.42 3.44	2.30 3.22	2.22 3.05	2.14 2.92	2.09 2.82	2.04 2.73	2.00 2.66	1.97 2.60	1.91 2.50	1.87 2.42	1.80 2.30	1.75 2.22	1.71 2.13	1.65 2.04	1.62 1.98	1.57 1.90	1.54 1.86	1.51 1.80	1.48 1.76	1.46 1.72	46
48	4.04 7.19	3.19 5.08	2.80 4.22	2.56 3.74	2.41 3.42	2.30 3.20	2.21 3.04	2.14 2.90	2.08 2.80	2.03 2.71	1.99 2.64	1.96 2.58	1.90 2.48	1.86 2.40	1.79 2.28	1.74 2.20	1.70 2.11	1.64 2.02	1.61 1.96	1.56 1.88	1.53 1.84	1.50 1.78	1.47 1.73	1.45 1.70	48
50	4.03 7.17	3.18 4.06	2.79 4.20	2.56 3.72	2.40 3.41	2.29 3.18	2.20 3.02	2.13 2.88	2.07 2.78	2.02 2.70	1.98 2.62	1.95 2.56	1.90 2.46	1.85 2.39	1.78 2.26	1.74 2.18	1.69 2.10	1.63 2.00	1.60 1.94	1.55 1.86	1.52 1.82	1.48 1.76	1.46 1.71	1.44 1.68	50
55	4.02 7.12	3.17 5.01	2.78 4.16	2.54 3.68	2.38 3.37	2.27 3.15	2.18 2.98	2.11 2.85	2.05 2.75	2.00 2.66	1.97 2.59	1.93 2.53	1.88 2.43	1.83 2.35	1.76 2.23	1.72 2.15	1.67 2.06	1.61 1.96	1.58 1.90	1.52 1.82	1.50 1.78	1.46 1.71	1.43 1.66	1.41 1.64	55
60	4.00 7.08	3.15 4.98	2.76 4.13	2.52 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63	1.95 2.56	1.92 2.50	1.86 2.40	1.81 2.32	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.93	1.56 1.87	1.50 1.79	1.48 1.74	1.44 1.68	1.41 1.63	1.39 1.60	60
65	3.99 7.04	3.14 4.95	2.75 4.10	2.51 3.62	2.36 3.31	2.24 3.09	2.15 2.93	2.08 2.79	2.02 2.70	1.98 2.61	1.94 2.54	1.90 2.47	1.85 2.37	1.80 2.30	1.73 2.18	1.68 2.09	1.63 2.00	1.57 1.90	1.54 1.84	1.49 1.84	1.46 1.76	1.42 1.71	1.39 1.60	1.37 1.56	65
70	3.98 7.01	3.13 4.92	2.74 4.08	2.50 3.60	2.35 3.29	2.23 3.07	2.14 2.91	2.07 2.77	2.01 2.67	1.97 2.59	1.93 2.51	1.89 2.45	1.84 2.35	1.79 2.28	1.72 2.15	1.67 2.07	1.62 1.98	1.56 1.88	1.53 1.82	1.47 1.74	1.45 1.69	1.40 1.62	1.37 1.56	1.35 1.53	70
80	3.96 6.96	3.11 4.88	2.72 4.04	2.48 3.56	2.33 3.25	2.21 3.04	2.12 2.87	2.05 2.74	1.99 2.64	1.95 2.55	1.91 2.48	1.88 2.41	1.82 2.32	1.77 2.24	1.70 2.11	1.65 2.03	1.60 1.94	1.54 1.84	1.51 1.78	1.45 1.70	1.42 1.65	1.38 1.57	1.35 1.52	1.32 1.49	80
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.30 3.20	2.19 2.99	2.10 2.82	2.03 2.69	1.97 2.59	1.92 2.51	1.88 2.43	1.85 2.36	1.79 2.26	1.75 2.19	1.68 2.06	1.63 1.98	1.57 1.89	1.51 1.79	1.48 1.73	1.42 1.64	1.39 1.59	1.34 1.51	1.30 1.46	1.28 1.43	100
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.08 2.79	2.01 2.65	1.95 2.56	1.90 2.47	1.86 2.40	1.83 2.33	1.77 2.23	1.72 2.15	1.65 2.03	1.60 1.94	1.55 1.85	1.49 1.75	1.45 1.68	1.39 1.59	1.36 1.54	1.31 1.46	1.27 1.40	1.25 1.37	125
150	3.91 6.81	3.06 4.75	2.67 3.91	2.43 3.44	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.62	1.94 2.53	1.89 2.44	1.85 2.37	1.82 2.30	1.76 2.20	1.71 2.12	1.64 2.00	1.59 1.91	1.54 1.83	1.47 1.72	1.44 1.66	1.37 1.56	1.34 1.51	1.29 1.43	1.25 1.37	1.22 1.33	150
200	3.89 6.76	3.04 4.71	2.65 3.88	2.41 3.41	2.26 3.11	2.14 2.90	2.05 2.73	1.98 2.60	1.92 2.50	1.87 2.41	1.83 2.34	1.80 2.28	1.74 2.17	1.69 2.09	1.62 1.97	1.57 1.88	1.52 1.79	1.45 1.69	1.42 1.62	1.35 1.53	1.32 1.48	1.26 1.39	1.22 1.33	1.19 1.28	200
400	3.86 6.70	3.02 4.66	2.62 3.83	2.39 3.36	2.23 3.06	2.12 2.85	2.03 2.69	1.96 2.55	1.90 2.46	1.85 2.37	1.81 2.29	1.78 2.23	1.72 2.12	1.67 2.04	1.60 1.92	1.54 1.84	1.49 1.74	1.42 1.64	1.38 1.57	1.32 1.47	1.28 1.42	1.22 1.32	1.16 1.24	1.13 1.19	400
1000	3.85 6.66	3.00 4.62	2.61 2.80	2.38 3.34	2.22 3.04	2.10 2.82	2.02 2.66	1.95 2.53	1.89 2.43	1.84 2.34	1.80 2.26	1.76 2.20	1.70 2.09	1.65 2.01	1.58 1.89	1.53 1.81	1.47 1.71	1.41 1.61	1.36 1.54	1.30 1.44	1.26 1.38	1.19 1.28	1.13 1.19	1.08 1.11	1000
∞	3.84 6.64	2.99 4.60	2.60 3.78	2.37 3.32	2.21 3.02	2.09 2.80	2.01 2.64	1.94 2.51	1.88 2.41	1.83 2.32	1.79 2.24	1.75 2.18	1.69 2.07	1.64 1.99	1.57 1.87	1.52 1.79	1.46 1.69	1.40 1.59	1.35 1.52	1.28 1.41	1.24 1.36	1.17 1.25	1.11 1.15	1.00 1.00	∞

Significant Studentized Ranges for 5% Level New Multiple Range test

Error df	p = number of means for range being tested													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	2.46	3.47
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
24	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47

Significant Studentized Ranges for 1% Level New Multiple Range test

Error df	p = number of means for range being tested														
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	
1	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	
2	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	
3	8.26	8.50	8.60	8.70	8.80	8.90	8.95	9.00	9.03	9.06	9.10	9.20	9.30	9.30	
4	6.51	6.80	6.90	7.00	7.10	7.15	7.20	7.25	7.30	7.35	7.40	7.45	7.50	7.50	
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.50	6.60	6.60	6.70	6.75	6.80	
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.05	6.10	6.20	6.25	6.30	6.30	
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.80	5.85	5.90	5.95	6.00	6.00	
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.50	5.60	5.70	5.75	5.80	5.80	
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.40	5.50	5.55	5.60	5.70	5.70	
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55	
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39	
12	4.32	4.55	4.68	4.76	4.81	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26	
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15	
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	3.87	4.91	4.96	5.00	5.04	5.06	5.07	
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00	
16	4.13	4.34	4.45	4.54	4.60	4.67	4.7	4.76	4.79	4.84	4.88	4.91	4.93	4.94	
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89	
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85	
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82	
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79	
22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75	
24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72	
26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69	
28	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67	
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65	
40	3.82	3.99	4.10	4.17	4.21	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59	
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53	
100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48	
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41	

Student's t-distribution

n	.9	.8	.7	.6	.5	.4	.3	.2	.1	.05	.02	.01	.001
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	.130	.263	.402	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	.128	.259	.394	.538	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	.128	.258	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	.127	.257	.391	.532	.686	.859	1.063	1.323	1.72641	2.080	2.518	2.831	3.819
22	.127	.256	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	.127	.256	.390	.532	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	.127	.256	.390	.531	.685	.857	1.059	1.318	1.711	2.066	2.492	2.797	3.745
25	.127	.256	.390	.531	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	.127	.256	.390	.531	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	.127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	.127	.256	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	.127	.256	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	.126	.255	.388	.529	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	.126	.254	.387	.527	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

ARCSINE TRANSFORMATION

%	0	1	2	3	4	5	6	7	8	9
0.0	0.00	0.57	0.81	0.99	1.15	1.28	1.40	1.52	1.62	1.72
0.1	1.81	1.90	1.99	2.07	2.14	2.22	2.29	2.36	2.43	2.50
0.2	2.56	2.63	2.69	2.75	2.81	2.87	2.92	2.98	3.03	3.09
0.3	3.14	3.19	3.24	3.29	3.34	3.39	3.44	3.49	3.53	3.58
0.4	3.63	3.67	3.72	3.76	3.80	3.85	3.89	3.93	3.97	4.01
0.5	4.05	4.10	4.14	4.17	4.21	4.25	4.29	4.33	4.37	4.41
0.6	4.44	4.48	4.52	4.55	4.59	4.62	4.66	4.70	4.73	4.76
0.7	4.80	4.83	4.87	4.90	4.93	4.97	5.00	5.03	5.07	5.10
0.8	5.13	5.16	5.20	5.23	5.26	5.29	5.32	5.35	5.38	5.41
0.9	5.44	5.47	5.50	5.53	5.56	5.59	5.62	5.65	5.68	5.71
1	5.74	6.02	6.29	6.55	6.80	7.03	7.27	7.49	7.71	7.92
2	8.13	8.33	8.53	8.72	8.91	9.10	9.28	9.46	9.63	9.80
3	9.97	10.14	10.30	10.47	10.63	10.78	10.94	11.09	11.24	11.39
4	11.54	11.68	11.83	11.97	12.11	12.25	12.38	12.52	12.66	12.79
5	12.92	13.05	13.18	13.31	13.44	13.56	13.69	13.81	13.94	14.06
6	14.18	14.30	14.42	14.54	14.65	14.77	14.89	15.00	15.12	15.23
7	15.34	15.45	15.56	15.68	15.79	15.89	16.00	16.11	16.22	16.32
8	16.43	16.54	16.64	16.74	16.85	16.95	17.05	17.15	17.26	17.36
9	17.46	17.56	17.66	17.76	17.85	17.95	18.05	18.15	18.24	18.34
10	18.43	18.53	18.63	18.72	18.81	18.91	19.00	19.09	19.19	19.28
11	19.37	19.46	19.55	19.64	19.73	19.82	19.91	20.00	20.09	20.18
12	20.27	20.36	20.44	20.53	20.62	20.70	20.79	20.88	20.96	21.05
13	21.13	21.22	21.30	21.39	21.47	21.56	21.64	21.72	21.81	21.89
14	21.97	22.06	22.14	22.22	22.30	22.38	22.46	22.54	22.63	22.71
15	22.79	22.87	22.95	23.03	23.11	23.18	23.26	23.34	23.42	23.50
16	23.58	23.66	23.73	23.81	23.89	23.97	24.04	24.12	24.20	24.27
17	24.35	24.43	24.50	24.58	24.65	24.73	24.80	24.88	24.95	25.03
18	25.10	25.18	25.25	25.33	25.40	25.47	25.55	25.62	25.70	25.77
19	25.84	25.91	25.99	26.06	26.13	26.21	26.28	26.35	26.42	26.49
20	26.57	26.64	26.71	26.78	26.85	26.92	26.99	27.06	27.13	27.20
21	27.27	27.35	27.42	27.49	27.56	27.62	27.69	27.76	27.83	27.90
22	27.97	28.04	28.11	28.18	28.25	28.32	28.39	28.45	28.52	28.59
23	28.66	28.73	28.79	28.86	28.93	29.00	29.06	29.13	29.20	29.27
24	29.33	29.40	29.47	29.53	29.60	29.67	29.73	29.80	29.87	29.93
25	30.00	30.07	30.13	30.20	30.26	30.33	30.40	30.46	30.53	30.59
26	30.66	30.72	30.79	30.85	30.92	30.98	31.05	31.11	31.18	31.24
27	31.31	31.37	31.44	31.50	31.56	31.63	31.69	31.76	31.82	31.88
28	31.95	32.01	32.08	32.14	32.20	32.27	32.33	32.39	32.46	32.52
29	32.58	32.65	32.71	32.77	32.83	32.90	32.96	33.02	33.09	33.15
30	33.21	33.27	33.34	33.40	33.46	33.52	33.58	33.65	33.71	33.77
31	33.83	33.90	33.96	34.02	34.08	34.14	34.20	34.27	34.33	34.39
32	34.45	34.51	34.57	34.63	34.70	34.76	34.82	34.88	34.94	35.00
33	35.06	35.12	35.18	35.24	35.30	35.37	35.43	35.49	35.55	35.61
34	35.67	35.73	35.79	35.85	35.91	35.97	36.03	36.09	36.15	36.21
35	36.27	36.33	36.39	36.45	36.51	36.57	36.63	36.69	36.75	36.81
36	36.87	36.93	36.99	37.05	37.11	37.17	37.23	37.29	37.35	37.41
37	37.46	37.52	37.58	37.64	37.70	37.76	37.82	37.88	37.94	38.00
38	38.06	38.12	38.17	38.23	38.29	38.35	38.41	38.47	38.53	38.59
39	38.65	38.70	38.76	38.82	38.88	38.94	39.00	39.06	39.11	39.17

%	0	1	2	3	4	5	6	7	8	9
40	39.23	39.29	39.35	39.41	39.47	39.52	39.58	39.64	39.70	39.76
41	39.82	39.87	39.93	39.99	40.05	40.11	40.16	40.22	40.28	40.34
42	40.40	40.45	40.51	40.57	40.63	40.69	40.74	40.80	40.86	40.92
43	40.98	41.03	41.09	41.15	41.21	41.27	41.32	41.38	41.44	41.50
44	41.55	41.61	41.67	41.73	41.78	41.84	41.90	41.96	42.02	42.07
45	42.13	42.19	42.25	42.30	42.36	42.42	42.48	42.53	42.59	42.65
46	42.71	42.76	42.82	42.88	42.94	42.99	43.05	43.11	43.17	43.22
47	43.28	43.34	43.39	43.45	43.51	43.57	43.62	43.68	43.74	43.80
48	43.85	43.91	43.97	44.03	44.08	44.14	44.20	44.26	44.31	44.37
49	44.43	44.48	44.54	44.60	44.66	44.71	44.77	44.83	44.89	44.94
50	45.00	45.06	45.11	45.17	45.23	45.29	45.34	45.40	45.46	45.52
51	45.57	45.63	45.69	45.74	45.80	45.86	45.92	45.97	46.03	46.09
52	46.15	46.20	46.26	46.32	46.38	46.43	46.49	46.55	46.61	46.66
53	46.72	46.78	46.83	46.89	46.95	47.01	47.06	47.12	47.18	47.24
54	47.29	47.35	47.41	47.47	47.52	47.58	47.64	47.70	47.75	47.81
55	47.87	47.93	47.98	48.04	48.10	48.16	48.22	48.27	48.33	48.39
56	48.45	48.50	48.56	48.62	48.68	48.73	48.79	48.85	48.91	48.97
57	49.02	49.08	49.14	49.20	49.26	49.31	49.37	49.43	49.49	49.55
58	49.60	49.66	49.72	49.78	49.84	49.89	49.95	50.01	50.07	50.13
59	50.18	50.24	50.30	50.36	50.42	50.48	50.53	50.59	50.65	50.71
60	50.77	50.83	50.89	50.94	51.00	51.06	51.12	51.18	51.24	51.30
61	51.35	51.41	51.47	51.53	51.59	51.65	51.71	51.77	51.83	51.88
62	51.94	52.00	52.06	52.12	52.18	52.24	52.30	52.36	52.42	52.48
63	52.54	52.59	52.65	52.71	52.77	52.83	52.89	52.95	53.01	53.07
64	53.13	53.19	53.25	53.31	53.37	53.43	53.49	53.55	53.61	53.67
65	53.73	53.79	53.85	53.91	53.97	54.03	54.09	54.15	54.21	54.27
66	54.33	54.39	54.45	54.51	54.57	54.63	54.70	54.76	54.82	54.88
67	54.94	55.00	55.06	55.12	55.18	55.24	55.30	55.37	55.43	55.49
68	55.55	55.61	55.67	55.73	55.80	55.86	55.92	55.98	56.04	56.10
69	56.17	56.23	56.29	56.35	56.42	56.48	56.54	56.60	56.66	56.73
70	56.79	56.85	56.91	56.98	57.04	57.10	57.17	57.23	57.29	57.35
71	57.42	57.48	57.54	57.61	57.67	57.73	57.80	57.86	57.92	57.99
72	58.05	58.12	58.18	58.24	58.31	58.37	58.44	58.50	58.56	58.63
73	58.69	58.76	58.82	58.89	58.95	59.02	59.08	59.15	59.21	59.28
74	59.34	59.41	59.47	59.54	59.60	59.67	59.74	59.80	59.87	59.93
75	60.00	60.07	60.13	60.20	60.27	60.33	60.40	60.47	60.53	60.60
76	60.67	60.73	60.80	60.87	60.94	61.00	61.07	61.14	61.21	61.27
77	61.34	61.41	61.48	61.55	61.61	61.68	61.75	61.82	61.89	61.96
78	62.03	62.10	62.17	62.24	62.31	62.38	62.44	62.51	62.58	62.65
79	62.73	62.80	62.87	62.94	63.01	63.08	63.15	63.22	63.29	63.36
80	63.43	63.51	63.58	63.65	63.72	63.79	63.87	63.94	64.01	64.09
81	64.16	64.23	64.30	64.38	64.45	64.53	64.60	64.67	64.75	64.82
82	64.90	64.97	65.05	65.12	65.20	65.27	65.35	65.42	65.50	65.57
83	65.65	65.73	65.80	65.88	65.96	66.03	66.11	66.19	66.27	66.34
84	66.42	66.50	66.58	66.66	66.74	66.82	66.89	66.97	67.05	67.13
85	67.21	67.29	67.37	67.46	67.54	67.62	67.70	67.78	67.86	67.94
86	68.03	68.11	68.19	68.28	68.36	68.44	68.53	68.61	68.70	68.78
87	68.87	68.95	69.04	69.12	69.21	69.30	69.38	69.47	69.56	69.64
88	69.73	69.82	69.91	70.00	70.09	70.18	70.27	70.36	70.45	70.54
89	70.63	70.72	70.81	70.91	71.00	71.09	71.19	71.28	71.37	71.47

%	0	1	2	3	4	5	6	7	8	9
90	71.57	71.66	71.76	71.85	71.95	72.05	72.15	72.24	72.34	72.44
91	72.54	72.64	72.74	72.85	72.95	73.05	73.15	73.26	73.36	73.46
92	73.57	73.68	73.78	73.89	74.00	74.11	74.21	74.32	74.44	74.55
93	74.66	74.77	74.88	75.00	75.11	75.23	75.35	75.46	75.58	75.70
94	75.82	75.94	76.06	76.19	76.31	76.44	76.56	76.69	76.82	76.95
95	77.08	77.21	77.34	77.48	77.62	77.75	77.89	78.03	78.17	78.32
96	78.46	78.61	78.76	78.91	79.06	79.22	79.37	79.53	79.70	79.86
97	80.03	80.20	80.37	80.54	80.72	80.90	81.09	81.28	81.47	81.67
98	81.87	82.08	82.29	82.51	82.73	82.97	83.20	83.45	83.71	83.98
99.0	84.26	84.29	84.32	84.35	84.38	84.41	84.44	84.47	84.50	84.53
99.1	84.56	84.59	84.62	84.65	84.68	84.71	84.74	84.77	84.80	84.84
99.2	84.87	84.90	84.93	84.97	85.00	85.03	85.07	85.10	85.13	85.17
99.3	85.20	85.24	85.27	85.30	85.34	85.38	85.41	85.45	85.48	85.52
99.4	85.56	85.59	85.63	85.67	85.71	85.75	85.79	85.83	85.86	85.90
99.5	85.95	85.99	86.03	86.07	86.11	86.15	86.20	86.24	86.28	86.33
99.6	86.37	86.42	86.47	86.51	86.56	86.61	86.66	86.71	86.76	86.81
99.7	86.86	86.91	86.97	87.02	87.08	87.13	87.19	87.25	87.31	87.37
99.8	87.44	87.50	87.57	87.64	87.71	87.78	87.86	87.93	88.01	88.10
99.9	88.19	88.28	88.38	88.48	88.60	88.72	88.85	89.01	89.19	89.43
100.0	90.00									